



Eixo 1 Educação - Descritivo do Projeto

Sistema Inteligente para Detecção de Anomalias no Cartão PDAF

1. Contextualização

No âmbito do Plano de Trabalho do Convênio nº 01/2024- CIIADF, coordenado pelo Instituto Hardware Br – HBR foram desenvolvidas soluções inovadoras para modernização do ensino público no Distrito Federal, com foco em personalização da aprendizagem, apoio a educadores e uso ético da Inteligência Artificial.

No contexto da gestão educacional, o Programa de Descentralização Administrativa e Financeira (PDAF) desempenha papel estratégico ao repassar recursos diretamente às escolas públicas, promovendo maior autonomia na execução de despesas. Entretanto, esse modelo descentralizado traz desafios relevantes relacionados ao controle, transparência e auditoria dos recursos públicos. Diante desse cenário, torna-se essencial o desenvolvimento de mecanismos automatizados que apoiem a identificação de inconsistências, irregularidades e possíveis fraudes nas transações realizadas por meio do Cartão PDAF.

2. Objetivo do Projeto

Este projeto tem como objetivo o desenvolvimento de uma solução baseada em Inteligência Artificial para detecção de anomalias em transações financeiras do Cartão PDAF, contribuindo para o aprimoramento dos processos de auditoria e controle de recursos públicos.

Como objetivos específicos, destacam-se:

- Analisar dados financeiros relacionados ao PDAF;
- Desenvolver modelos preditivos para identificação de padrões anômalos;
- Avaliar o desempenho de diferentes algoritmos de aprendizado de máquina;
- Propor uma arquitetura de solução escalável para evolução futura em nível de produto.

3. Metodologia

O projeto adotou uma abordagem de pesquisa exploratória, com foco na compreensão dos dados e na experimentação de técnicas de aprendizado de máquina.

As principais etapas metodológicas incluem:

- **Coleta de dados:** utilização de dados públicos do PDAF e datasets de referência com características semelhantes (ex.: transações financeiras com rotulação de fraude);
- **Análise exploratória:** identificação de padrões, outliers e correlações relevantes;
- **Preparação dos dados:** construção de pipeline para tratamento, balanceamento de classes e estruturação das variáveis;
- **Modelagem:** aplicação de diferentes algoritmos de aprendizado supervisionado e não supervisionado, incluindo:



- Random Forest
 - Isolation Forest
 - XGBoost
 - CatBoost
 - Modelos baseados em Transformers para dados tabulares
- **Avaliação:** comparação dos modelos com base em métricas como precisão, recall e F1-score.

Além disso, foram exploradas técnicas de balanceamento de dados (oversampling e undersampling) para lidar com o desbalanceamento típico de problemas de fraude.

4. Resultados e Contribuições

Os experimentos demonstraram que modelos supervisionados apresentaram melhor desempenho geral na detecção de anomalias, especialmente em termos de equilíbrio entre precisão e recall.

Destacam-se os seguintes resultados:

- O modelo Random Forest apresentou desempenho robusto, com bom equilíbrio entre identificação de fraudes e redução de falsos positivos;
- O modelo XGBoost apresentou alta capacidade de generalização, com métricas consistentes mesmo em cenários desbalanceados;
- O modelo Isolation Forest apresentou alta sensibilidade (recall), porém com elevado número de falsos positivos;
- Modelos baseados em Transformers demonstraram potencial para capturar relações complexas entre variáveis, ainda que com maior custo computacional.

Adicionalmente, foi proposto um modelo híbrido, combinando técnicas supervisionadas e não supervisionadas, com o objetivo de otimizar a detecção de anomalias e reduzir o volume de transações a serem analisadas.

5. Evolução Esperada da PoC para Produto

A startup selecionada deverá evoluir a PoC de detecção de anomalias no Cartão PDAF para uma solução operacional demonstrável, capaz de apoiar auditoria, controle interno e priorização de casos suspeitos no contexto da Secretaria de Educação. Considerando o ciclo de 180 dias e o aporte equity-free de R\$ 100 mil, a expectativa é acelerar a adaptação de uma solução já existente ao blueprint da PoC, com foco em integração, validação, segurança, rastreabilidade e prontidão para contratação pública. O projeto original já demonstrou viabilidade técnica com uso de modelos supervisionados e não supervisionados, incluindo Random Forest, XGBoost, Isolation Forest e abordagens híbridas para identificação de padrões anômalos em transações financeiras do PDAF. Ao final dos 180 dias, a solução deverá ser capaz de ingerir dados reais ou amostras representativas



do PDAF, processar transações financeiras, identificar anomalias com critérios explicáveis e apresentar os casos priorizados em uma interface de apoio à auditoria. O resultado esperado não é apenas um modelo com boa métrica estatística, mas um sistema que ajude o gestor público a responder perguntas práticas: quais transações merecem análise? Por quê? Qual o grau de risco? Que evidências sustentam a classificação? Qual o próximo passo recomendado para a equipe de controle?

A evolução esperada deve contemplar, no mínimo:

i. Produto funcional aplicado ao fluxo de auditoria

A startup deverá entregar uma versão funcional da solução com painel de análise, fila de casos suspeitos, filtros por escola, período, natureza da despesa, fornecedor, valor, categoria e score de risco. A ferramenta deve permitir que usuários autorizados visualizem os principais sinais de anomalia, revisem transações individualmente e registrem decisões ou encaminhamentos.

ii. Motor de risco com explicabilidade mínima

Cada alerta gerado deverá apresentar justificativas compreensíveis para o auditor ou gestor público. Exemplos: valor atípico para determinada escola, frequência incomum de compras, fornecedor recorrente fora do padrão, fragmentação de despesas, concentração temporal, categoria incompatível ou desvio em relação ao histórico. A solução não pode operar como “caixa-preta”; precisa fornecer explicações auditáveis.

iii. Validação de desempenho em base representativa

A startup deverá validar o desempenho do modelo em base histórica, real ou simulada, comparando precisão, recall, taxa de falsos positivos e utilidade operacional dos alertas. O critério de sucesso não deve ser apenas maximizar detecção, mas reduzir ruído e priorizar casos relevantes. Em contexto público, um modelo que gera alerta demais sem priorização útil vira custo operacional.

iv. Pipeline de dados documentado e replicável

A solução deverá entregar pipeline documentado para ingestão, tratamento, normalização, versionamento e atualização dos dados. Devem estar claros os campos necessários, origem dos dados, periodicidade de atualização, regras de limpeza, tratamento de inconsistências e limitações conhecidas.

v. Segurança, LGPD e trilha de auditoria

A solução deverá contemplar controle de acesso por perfil, logs de uso, registro de decisões, rastreabilidade dos alertas e diretrizes de tratamento de dados conforme LGPD. Mesmo quando os dados não forem sensíveis no sentido estrito, o uso de informações financeiras públicas exige governança, integridade e controle de acesso.



vi. Evidência de adoção operacional

A startup deverá realizar uma simulação ou piloto assistido com usuários da Secretaria ou equipe indicada pelo programa, demonstrando o uso da solução em um fluxo realista de auditoria. A entrega deve incluir feedback dos usuários, ajustes priorizados e evidências de que a ferramenta reduz tempo de triagem ou melhora a priorização de casos.

vii. Pacote de contratação e escala

Ao final do ciclo, a startup deverá entregar documentação para subsidiar contratação: descrição técnica da solução, arquitetura, requisitos de infraestrutura, integrações necessárias, matriz de responsabilidades, SLAs sugeridos, estimativa de custos, plano de implantação, riscos, limitações e roadmap de evolução. Essa documentação deve permitir que a Secretaria avalie a adoção da solução com menor ambiguidade técnica e menor risco de procurement.

Marcos sugeridos de evolução

- **Até 60 dias:** aderência ao blueprint validada, acesso ou simulação dos dados definida, arquitetura de integração aprovada, baseline de métricas estabelecido e primeira versão funcional do pipeline.
- **Até 120 dias:** motor de risco calibrado, painel operacional disponível, explicabilidade dos alertas implementada, testes com base representativa concluídos e validação intermediária com usuários.
- **Até 180 dias:** solução demonstrável em fluxo operacional, evidências de desempenho consolidadas, documentação técnica e operacional finalizada, plano de rollout preparado e pacote de contratação entregue.

Critério de chegada

A PoC será considerada evoluída com sucesso quando houver uma solução demonstrável, segura, documentada e operacionalmente útil para priorizar auditorias do PDAF, acompanhada de evidências suficientes para decidir sobre contratação, expansão ou novo ciclo de validação.

5. Considerações Finais

O projeto demonstra a viabilidade do uso de Inteligência Artificial para apoiar a fiscalização e a gestão de recursos públicos no âmbito educacional. Ao combinar análise de dados, modelagem preditiva e inovação tecnológica, a iniciativa contribui diretamente para o fortalecimento da governança e para a melhoria da qualidade dos serviços educacionais no Distrito Federal.