

Potencial de Projeto de Data Science e Inteligência Artificial

1. Contextualização dos Dados

O conjunto de dados disponibilizado é composto por três fontes principais:

- **cards_data.csv**: histórico detalhado de transações com cartões, incluindo atributos como valor, data, identificador do usuário, e código MCC (Merchant Category Code).
- **mcc_codes.json**: mapeamento completo dos códigos MCC para descrições de categorias de estabelecimentos comerciais (alimentação, serviços, transporte, saúde, etc.)
- **users_data.csv**: informações demográficas e comportamentais dos usuários, tais como perfil socioeconômico, localização geográfica e segmento de mercado.

Essas bases podem ser integradas para gerar insights profundos sobre padrões de consumo, comportamento de subscrição e riscos associados a fraudes.

2. Objetivos do Projeto

1. **Segmentação de Clientes**: agrupar usuários por perfil de consumo e preferências, permitindo campanhas de marketing mais assertivas.
2. **Deteção e Prevenção de Fraudes**: identificar padrões atípicos em tempo real, reduzindo prejuízos e aumentando a confiança no sistema.
3. **Recomendação de Produtos e Serviços**: sugerir ofertas personalizadas com base em histórico de gastos e categorias MCC.
4. **Análise de Churn**: prever probabilidade de cancelamento ou migração de serviços, antecipando ações de retenção.
5. **Otimização de Fluxo de Caixa**: prever volumes de transações futuros para gestão financeira e de liquidez.

3. Metodologia

1. **Aquisição e Integração**:
 2. Importação de arquivos CSV/JSON e junção por identificadores de usuário e código MCC.
 3. Criação de dicionários de tradução de códigos MCC para categorias legíveis.
4. **Pré-processamento**:
 5. Limpeza de valores ausentes e inconsistentes.
 6. Normalização e agregação temporal (diária, semanal, mensal).
7. **Análise Exploratória (EDA)**:
 8. Estatísticas descritivas (médias, medianas, desvios-padrão).
 9. Visualização de distribuições de gastos por categoria e região.
10. **Modelagem Preditiva**:
 11. **Clustering** (K-means, DBSCAN) para segmentação de clientes.
 12. **Classificação** (Random Forest, XGBoost) para detecção de fraudes.
 13. **Regressão** (LightGBM, Redes Neurais) para previsão de volume de transações.
 14. **Sistemas de Recomendação** (Filtragem Colaborativa e Baseada em Conteúdo) para ofertas.
15. **Validação e Avaliação**:
 16. Cruzamento de validação, métricas de AUC, precisão, recall e F1-Score.
 17. Testes A/B para mensurar impacto em campanhas.
18. **Deploy e Monitoramento**:
 19. Criação de pipelines em cloud (AWS/GCP) e monitoramento contínuo de performance.

4. Casos de Uso Práticos

- **Campanhas de Marketing Direcionado:** envio de cupons de desconto para usuários identificados com alto potencial de conversão em categorias específicas.
- **Antifraude em Tempo Real:** bloqueio automático de transações suspeitas, com alertas por SMS e e-mail.
- **Recomendação Inteligente:** dashboard para gerentes de produto sugerir pacotes de serviços baseados no comportamento de consumo.
- **Gestão de Carteira de Crédito:** previsão de inadimplência e níveis de risco por segmento.

5. Benefícios Esperados

- Aumento de **até 15%** na taxa de conversão de ofertas personalizadas.
- Redução de **20–30%** nos custos com fraudes e chargebacks.
- Melhoria de **10%** na retenção de clientes graças à análise de churn.
- Otimização financeira com projeções acuradas de fluxo de caixa.

6. Expansão com IA Avançada

- **Chatbots e Assistentes Virtuais:** integração com modelos de linguagem (LLMs) para atendimento inteligente.
- **Análise de Sentimento:** processamento de feedbacks e redes sociais para avaliar percepção da marca.
- **Visão Computacional:** reconhecimento de notas fiscais e comprovantes de pagamento para automação de conciliação financeira.

7. Roadmap e Próximos Passos

1. **Fase 1 (1–2 meses):** integração de dados, EDA e prototipagem de clustering.
2. **Fase 2 (3–4 meses):** desenvolvimento de modelos de fraude e recomendação, testes A/B.
3. **Fase 3 (5–6 meses):** deploy em produção, dashboards interativos e automação de relatórios.
4. **Governança:** definição de políticas de ética, privacidade e segurança dos dados.

8. Conclusão

A aplicação de técnicas de Data Science e Inteligência Artificial sobre o histórico de transações, categorias de estabelecimentos e perfil de usuários permite não apenas gerar valor imediato através de redução de riscos e aumento de receita, mas também instaurar uma cultura de decisão baseada em dados, garantindo vantagem competitiva sustentável.