

Transações Fraudulentas

Modelagem e Avaliação de Dados.

Matheus Gabardo Messias

Orientador: Professor Douglas Rocha Mendes

Análise e Desenvolvimento de Sistemas - Faculdade Estácio de Curitiba



Objetivos

- Dominar o ambiente de desenvolvimento e SVM.
- Problematização conforme os dados colhidos.
- Solução.
- Explorar funções de modelagem de dados.
- Explorar com percentual de base de testes.
- Exibir dados em gráficos utilizando alguma biblioteca de exibição de dados.
- Com base no dataset escolhido, aplicar ao modelo os algoritmos SVC, KNN e Árvore.
- Avaliar a diferença da Acurácia entre os percentuais de testes.

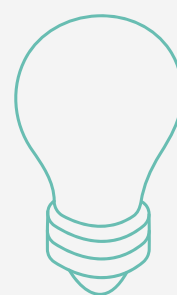




Dataset

Conjunto de dados

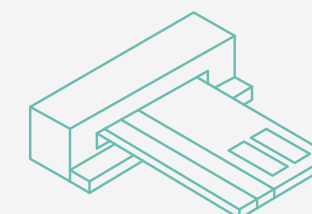
Os dados foram coletados a partir de transações online registradas por uma Fintech sediada nos Estados Unidos durante o período de 21/06/2020 a 28/06/2020. Fonte: Kaggle



Google Colab

Ambiente de desenvolvimento

Google Colab É um ambiente de notebooks que permite a criação de documentos interativos contendo códigos, textos formatados, imagens, gráficos e equações matemáticas. Sendo executado na nuvem do Google, não requer configuração.



SVM

Support Vector Machine




SVM É um algoritmo de aprendizado de máquina supervisionado que pode ser usado para desafios de classificação ou regressão. Seu foco principal é no treinamento e classificação de um dataset.



Bibliotecas



São ferramentas fundamentais para todo programador que trabalha com Python. Bibliotecas são conjuntos de módulos que contêm funções e métodos que facilitam a execução de tarefas específicas, como gerenciamento de dados, criação de gráficos, interação com bancos de dados, entre muitas outras funcionalidades.

| | |
|---|---|
| <div>Pandas</div> <div></div> | <p>Pandas é uma biblioteca de Python amplamente utilizada para manipulação e análise de dados de forma eficiente.</p> <p>Com ela podemos ler e escrever dados em vários formatos, realiza operações de limpeza, transformação e agregação de dados.</p> |
| <div>Numpy</div> <div></div> | <p>Fundamental para computação numérica, fornece estruturas de dados eficientes para armazenar e manipular arrays multidimensionais.</p> <p>Inclui funções para operações matemáticas avançadas em arrays, como álgebra linear, transformada de Fourier, geração de números aleatórios, entre outras.</p> |
| <div>matplotlib</div> <div></div> | <p>É especialmente útil para visualização dos dados complexos, oferecendo controle detalhado sobre a aparência dos gráficos.</p> |



Transações Fraudulentas

Segundo a Associação Brasileira de Comércio Eletrônico (ABComm), um e-commerce brasileiro de grande porte pode registrar tentativas de fraude a cada cinco segundos.

Danos

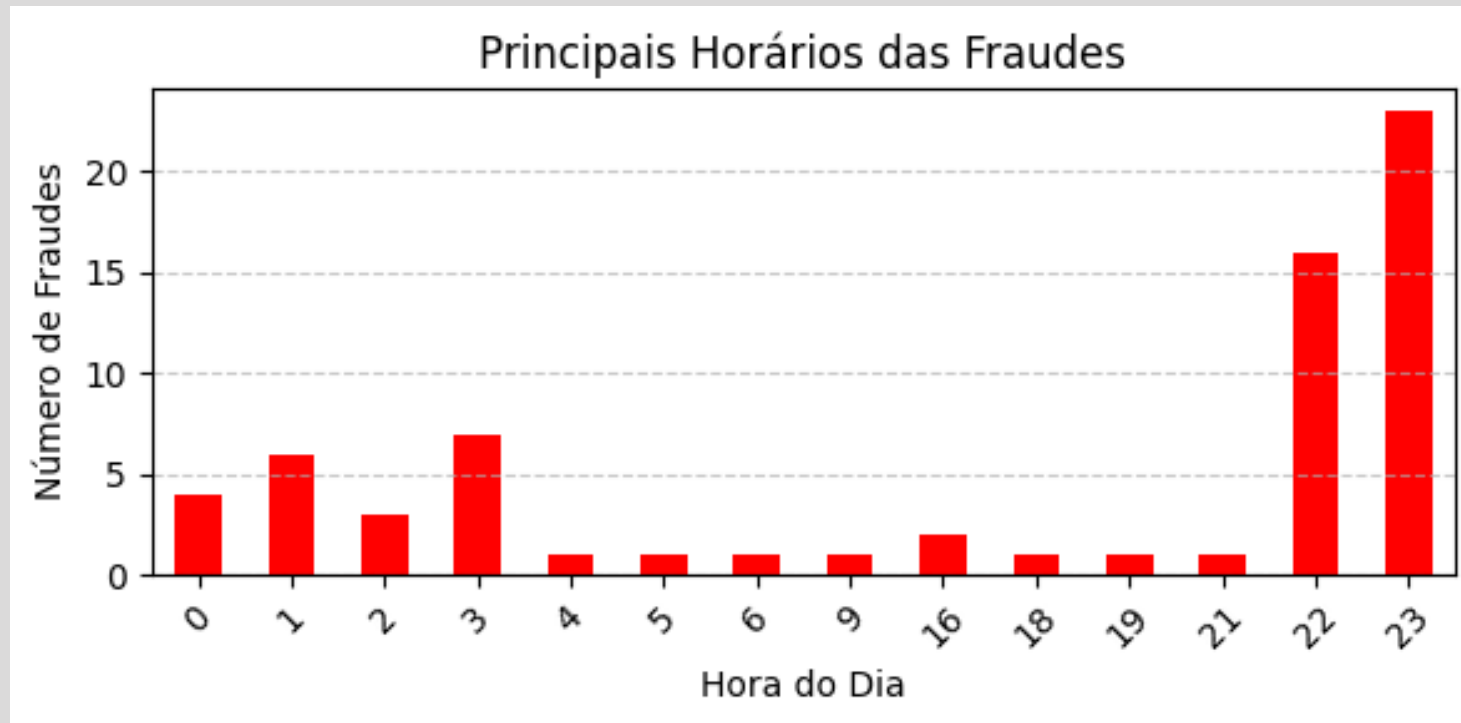
As consequências das fraudes de cartão de crédito podem incluir perda financeira para o titular, danos à sua reputação de crédito e a necessidade de passar por um processo de disputa com a instituição financeira para reverter as transações fraudulentas.

Como evitar

- Mantenha seus dados seguros.
- Use sites seguros.
- Proteja seu cartão físico.
- Fique atento a esquemas de phishing
 - Os golpistas muitas vezes se passam por instituições financeiras ou empresas legítimas para enganar as pessoas a fornecerem suas informações.

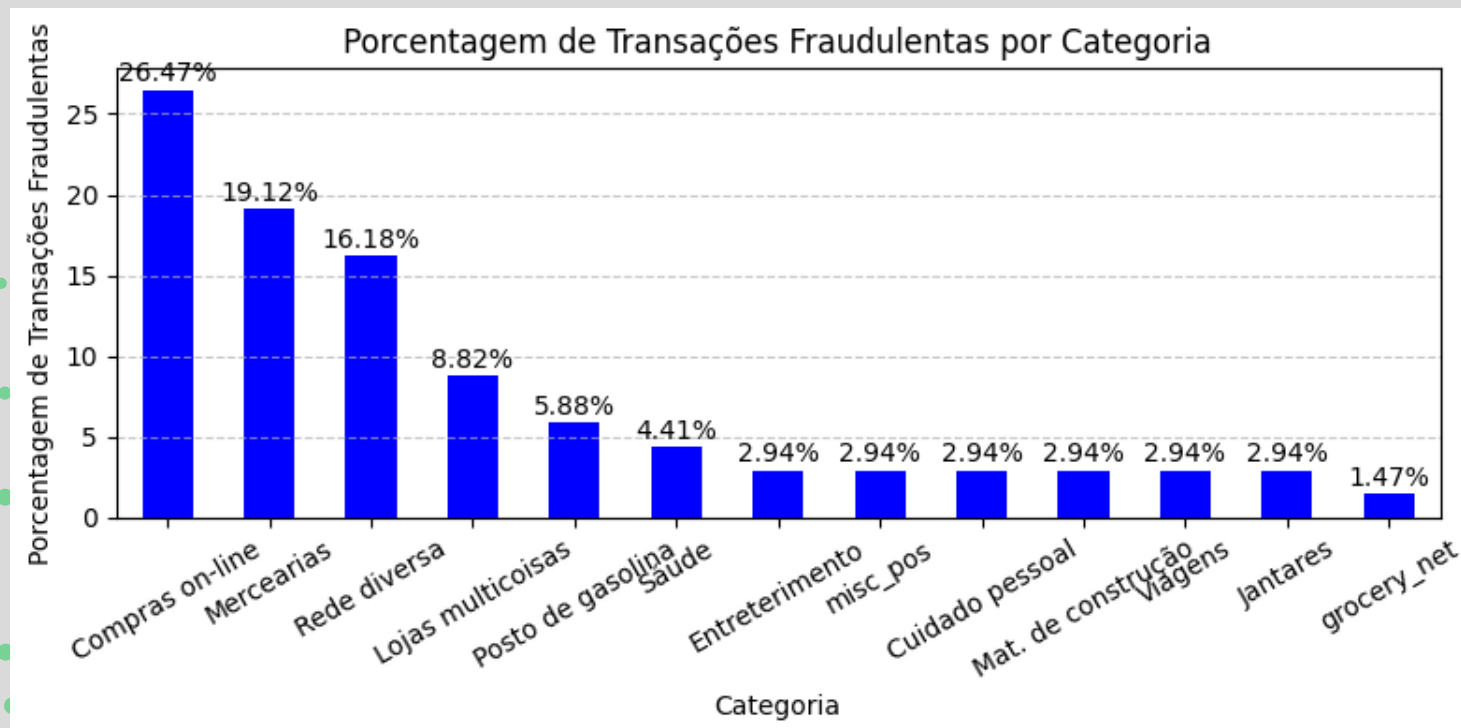


Gráficos gerados a partir do dataset escolhido



As fraudes de cartão de crédito podem ocorrer a qualquer hora do dia, mas pode haver uma percepção de que ocorrem mais frequentemente à noite por algumas razões:

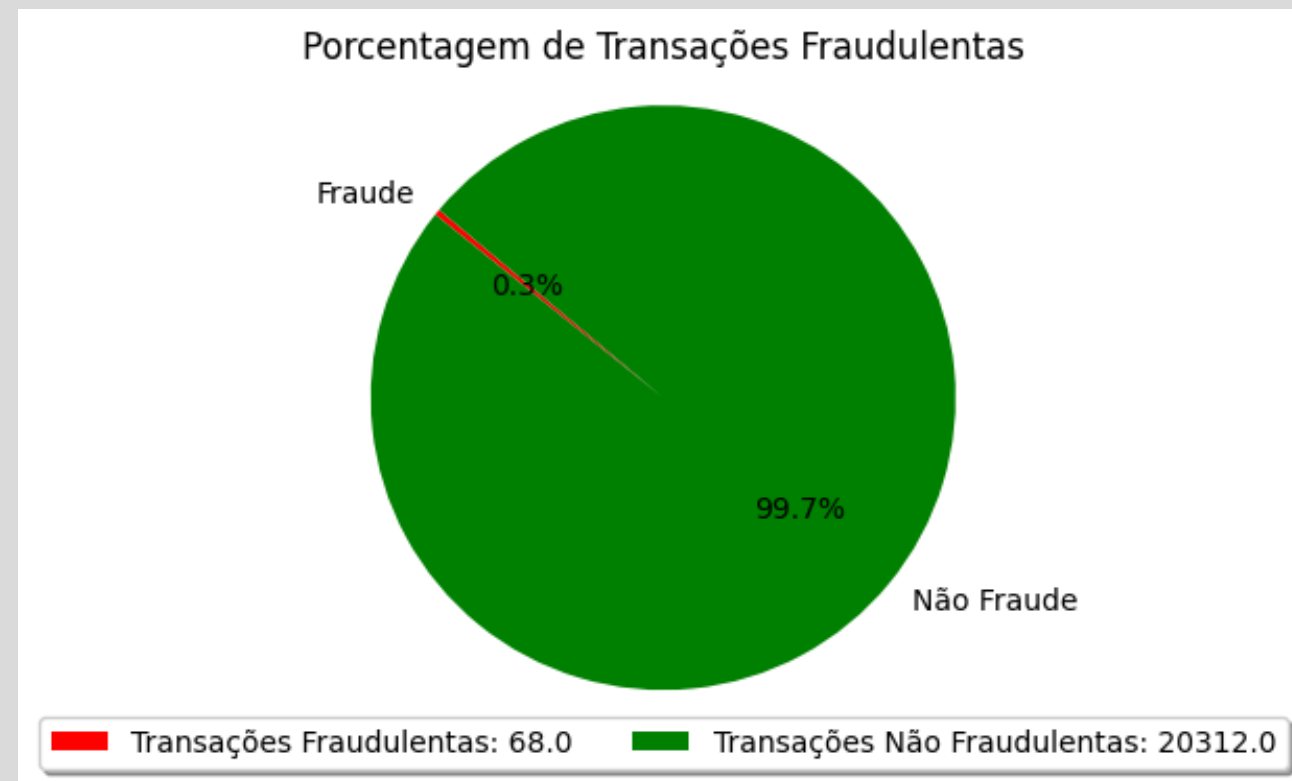
- Menor supervisão
- Maior atividade online
- Menos detecção imediata



As fraudes de cartão de crédito tendem a ocorrer com mais frequência em determinadas categorias de compras, incluindo:

- Posto de gasolina
 - Pode ser um alvo comum devido à natureza das transações rápidas e às vezes não supervisionadas.
- Compras online
 - A conveniência e a falta de interação pessoal podem tornar as compras online mais propensas a fraudes.
- Viagens
 - As transações relacionadas a viagens, como passagens aéreas e hospedagem, podem envolver grandes quantias, tornando-as alvos atraentes para fraudadores.

Outros



E o SVM?

No contexto da detecção de fraudes em transações de cartão de crédito, usamos o SVM para identificar padrões nos dados que separam transações legítimas de transações fraudulentas, mesmo quando os dados são altamente desbalanceados.

Dados utilizados para treinar os modelos;

- Valor da Transação (amt)
- Hora da Transação (hour)
- Categoria da Transação (category)
- Estado da Transação (state)
- Data de Nascimento (dob)

Modelos utilizados;

- SVC (Support Vector Classifier)
- KNN (K-Nearest Neighbors):
- Árvore de Decisão

- Esses modelos foram escolhidos por sua eficácia comprovada em problemas de classificação e sua capacidade de lidar com diferentes tipos de dados e padrões nos dados. Durante a fase de treinamento, cada modelo foi ajustado aos dados históricos para aprender a distinguir entre transações legítimas e fraudulentas.

- Ao considerar essas características e outras relevantes para o contexto específico do problema, os modelos de detecção podem ser treinados de forma mais eficaz para identificar e dados e probabilidades com maior precisão.



Avaliações

Avaliando os modelos 30%

- Acurácia do modelo SVC (30% de teste): 0.9955839057899902
- Acurácia do modelo KNN (30% de teste): 0.9952567877003599
- Acurácia do modelo Tree (30% de teste): 0.994438992476284

Avaliando os modelos 50%

- Acurácia do modelo SVC (50% de teste): 0.9964671246319922
- Acurácia do modelo KNN (50% de teste): 0.9960745829244357
- Acurácia do modelo Tree (50% de teste): 0.9965652600588812

Avaliando os modelos 50%

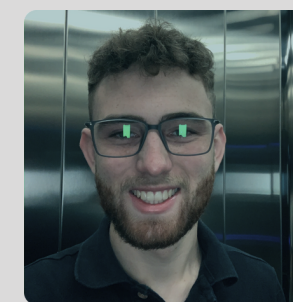
- Acurácia do modelo SVC (50% de teste): 0.9964671246319922
- Diferença de acurácia entre 30% e 50% para o modelo KNN: -0.000817795224075879
- Diferença de acurácia entre 30% e 50% para o modelo Tree: -0.0021262675825972632

Links;

[Notebook - Goggle colab](#)

[Dataset - Kaggle](#)

Obrigado!



Matheus Gabardo Messias

Desenvolvedor Frontend cursando o 5º período de Análise e desenvolvimento de sistemas na Faculdade Estácio de Curitiba.