**Análise Preditiva para Recuperação de Vendas: Relatório de Conclusão**

**1. Introdução e Objetivo**\*

Este projeto teve como objetivo principal desenvolver um modelo de análise preditiva para identificar clientes que, após abandonarem o carrinho de compras, possuem alta probabilidade de converter em uma venda. O foco da análise foi maximizar a *taxa de recuperação de vendas* por meio da identificação de *leads* qualificados.

**2. Metodologia**

O processo seguiu uma metodologia padrão de ciência de dados, garantindo a robustez e a confiabilidade dos resultados:

- **Preparação dos Dados**: Foram selecionadas features relevantes para a predição, como ‘preco’, ‘metodo\_pagamento’, ‘hora’ e ‘tipo\_plano’. Colunas que poderiam causar vazamento de dados, como ‘status\_compra’ e ‘taxa\_conversao’, foram removidas. Variáveis categóricas foram transformadas em formato numérico usando a técnica *One-Hot Encoding*.

- **Divisão e Treinamento:** O conjunto de dados foi dividido em 70% para treinamento e 30% para teste, garantindo que o desempenho do modelo fosse avaliado em dados não utilizados durante o aprendizado.

- **Modelagem Preditiva**: Foram testados dois algoritmos de aprendizado de máquina: *Random Forest e Regressão Logística.*

- **Avaliação**: A performance dos modelos foi avaliada com foco na métrica Recall, que mede a capacidade do modelo de identificar corretamente as oportunidades de venda.

**3. Análise e Resultados dos Modelos**

| Modelo Testado | Acurácia Geral | Recall (Conversão) | Principais Vantagens | Principais Desvantagens |
| --- | --- | --- | --- | --- |
| Random Forest | 78% | 0% | • Elevada acurácia geral  • Boa capacidade de generalização  • Robusto contra overfitting | • *Falhou completamente na identificação de oportunidades de conversão* |
| Regressão Logística | 49% | 50% | • *Identificou 50% das oportunidades reais de conversão*  • Modelo interpretável  • Adequado para identificação de leads qualificados | • Acurácia geral inferior  • Alto número de falsos positivos |

Apesar de o modelo de Random Forest ter apresentado uma acurácia superior (78%), sua performance para o objetivo do projeto foi nula, com um *Recall de 0%*. Isso significa que o modelo não conseguiu identificar corretamente nenhum cliente que, de fato, converteu, tornando-o ineficaz para a recuperação de vendas.

O modelo de *Regressão Logística*, por outro lado, demonstrou ser o mais adequado. Embora sua acurácia geral seja de 49%, ele conseguiu um *Recall de 50%*, o que indica que é capaz de encontrar metade de todos os clientes que converteram em vendas, mesmo que o modelo tenha gerado alguns "alarmes falsos" (Falsos Positivos).

Observação: Em contextos de vendas e marketing, o Recall é frequentemente mais importante que a acurácia geral, pois perder oportunidades de conversão tem impacto direto na receita.

**4. Conclusão e Recomendação Final**

Para o objetivo de recuperação de vendas, o modelo de Regressão Logística é a escolha ideal. Sua capacidade de identificar 50% das oportunidades perdidas representa um retorno tangível e direto para o negócio.

- Taxa Estimada de Recuperação de Vendas: A taxa estimada de recuperação é de 50% das vendas perdidas.

- Amostra Preditiva: O modelo gerou uma amostra com 1.829 clientes que, com base nos dados, apresentam alta probabilidade de conversão.

A recomendação é que essa amostra preditiva seja usada para direcionar os esforços de uma equipe de vendas ou de marketing, que pode focar seus recursos em clientes com maior probabilidade de retorno, otimizando o processo de recuperação de vendas.