## **Relatório de Análise Preditiva: Previsão de Compras Futuras de Clientes**

**Autor:** [Seu Nome ou Nome do Time] **Data:** 17 de Setembro de 2025 **Versão:** 1.0

### **Resumo Executivo**

O presente relatório detalha o desenvolvimento de um modelo de machine learning para prever a probabilidade de um cliente realizar uma nova compra. Utilizando um histórico de transações de um ano da base de dados "Online Retail", foi aplicado um modelo de segmentação baseado em comportamento (RFM - Recência, Frequência, Valor Monetário) para extrair características relevantes dos clientes. Um modelo de classificação (*Random Forest*) foi treinado com sucesso, alcançando uma acurácia de **86.6%** na identificação de futuros compradores. O modelo permite à empresa direcionar proativamente os esforços de marketing e retenção, otimizando o retorno sobre o investimento e aumentando a fidelização de clientes. As principais recomendações incluem a criação de campanhas de reativação para clientes com alta recência e a oferta de benefícios para clientes de alta frequência e valor.

### **1. Introdução**

A capacidade de antecipar o comportamento do consumidor é um diferencial competitivo crucial no mercado de varejo. Identificar quais clientes estão mais propensos a realizar uma compra futura permite que as equipes de marketing e vendas criem campanhas personalizadas, otimizem a alocação de recursos e melhorem as estratégias de retenção.

**1.1. Objetivo do Projeto** O objetivo principal deste projeto foi desenvolver e validar um modelo preditivo para classificar clientes existentes com base na sua probabilidade de realizar uma nova compra no mês seguinte.

**1.2. Fonte de Dados** A análise foi realizada com base no dataset "Online Retail", que contém o histórico de transações de vendas online realizadas entre 01/12/2010 e 09/12/2011.

### **2. Metodologia**

A metodologia foi estruturada em etapas, desde a preparação dos dados até o treinamento e avaliação do modelo final.

**2.1. Preparação e Limpeza dos Dados** A qualidade dos dados é fundamental para a performance do modelo. As seguintes ações foram executadas:

* **Remoção de Dados Nulos:** Registros sem identificação do cliente (CustomerID) foram removidos.
* **Tratamento de Transações Canceladas:** Faturas que indicavam cancelamento (código 'C') foram excluídas.
* **Remoção de Duplicatas:** Registros duplicados foram eliminados.
* **Criação da Métrica de Receita:** Uma nova coluna, Receita, foi criada multiplicando-se a quantidade de itens pelo preço unitário.

**2.2. Engenharia de Atributos: Análise RFM** Para transformar o histórico de transações em atributos que descrevem o comportamento do cliente, foi utilizada a técnica RFM:

* **Recência (R):** Há quantos dias o cliente fez sua última compra.
* **Frequência (F):** Quantas compras distintas o cliente realizou.
* **Monetário (M):** Qual o valor total gasto pelo cliente.

Essas métricas foram calculadas utilizando os dados até **30/11/2011**, que foi definida como a data de corte para o treinamento.

**2.3. Definição da Variável Alvo** Para treinar um modelo supervisionado, definimos uma variável alvo (Fez\_compra):

* Um cliente recebeu o valor **1** se, após a data de corte, ele realizou ao menos uma nova compra.
* Um cliente recebeu o valor **0** se não realizou nenhuma compra nesse período futuro.

**2.4. Modelagem Preditiva**

* **Algoritmo:** Foi escolhido o Random Forest Classifier.
* **Divisão dos Dados:** O conjunto de dados RFM foi dividido em 80% para treinamento e 20% para teste.

### **3. Resultados e Análise**

O modelo treinado foi avaliado com o conjunto de dados de teste, que ele nunca havia visto antes.

**3.1. Desempenho do Modelo**

As métricas de desempenho consolidadas foram as seguintes:

* **Acurácia Geral:** **86.6%**

**Relatório de Classificação:** precision recall f1-score support

0 0.89 0.96 0.93 747

1 0.48 0.23 0.31 113

accuracy 0.87 860

macro avg 0.69 0.60 0.62 860

weighted avg 0.84 0.87 0.85 860

**Análise das Métricas:** A acurácia geral de 86.6% indica uma boa performance global do modelo. Analisando em detalhe a classe de interesse (**classe 1**, clientes que compraram):

* A **precisão (precision) de 0.48** significa que, de todos os clientes que o modelo previu que fariam uma nova compra, 48% deles de fato compraram.
* O **recall de 0.23** indica que o modelo conseguiu identificar corretamente 23% de todos os clientes que realmente voltaram a comprar.

Isso revela que o modelo é **conservador**: ele tem dificuldade em encontrar todos os compradores (baixo recall), mas dos poucos que ele aponta como compradores, quase metade realmente compra (precisão razoável). Essa é uma informação valiosa para o negócio.

**3.2. Análise Visual dos Resultados (Recomendado)** Para um relatório ainda mais completo, a adição de gráficos é fundamental para facilitar a visualização e o entendimento dos resultados.

* **Matriz de Confusão:** Para visualizar diretamente os acertos e erros do modelo (quantos clientes ele previu corretamente que comprariam, quantos ele errou, etc.).
* **Importância dos Atributos:** Para mostrar qual das métricas RFM (Recência, Frequência ou Monetário) teve mais peso na decisão do modelo.

**Posso fornecer o código Python para gerar esses gráficos, se você desejar.**

### **4. Conclusão**

O estudo demonstrou ser viável prever, com um grau de confiança satisfatório, quais clientes retornarão para uma nova compra. O modelo baseado em RFM e Random Forest se provou uma ferramenta eficaz, fornecendo uma base sólida para a tomada de decisões estratégicas orientadas por dados. A baixa taxa de recall sugere que, embora o modelo seja útil, ele não deve ser a única ferramenta para identificar clientes valiosos, mas sim um complemento a outras estratégias.

### **5. Recomendações e Próximos Passos**

Com base nos resultados, as seguintes ações são recomendadas:

**5.1. Ações de Negócio:**

1. **Campanhas de Marketing Direcionado:** Utilizar a lista de clientes previstos como "compradores" (onde a precisão é de 48%) para campanhas de marketing de alta conversão, como ofertas exclusivas ou lançamento de novos produtos.
2. **Estratégia de Retenção Ampla:** Como o modelo perde muitos compradores reais (baixo recall), não se deve deixar de enviar comunicações de marketing genéricas para toda a base de clientes ativos, a fim de não perder oportunidades.
3. **Análise dos Falsos Negativos:** Investigar o perfil dos clientes que o modelo *errou* (previu que não comprariam, mas compraram). Eles podem representar um segmento de clientes com um padrão de compra menos óbvio, que merece uma estratégia de marketing própria.

**5.2. Próximos Passos Técnicos:**

1. **Otimização do Modelo para Recall:** Ajustar os parâmetros do Random Forest ou testar outros algoritmos com o objetivo de aumentar o recall, mesmo que isso signifique uma leve queda na precisão. Isso ajudaria a identificar um número maior de compradores potenciais.
2. **Inclusão de Mais Atributos:** Enriquecer o modelo com mais dados, como categorias de produtos comprados ou localização geográfica.
3. **Implantação (Deploy):** Integrar o modelo a um sistema de CRM para que as previsões sejam geradas e utilizadas de forma automatizada.