



UFOP

Mineração de Dados no E-commerce

Olist: Previsão de Satisfação, Agrupamento Logístico e Associação de Produtos

Beatriz Evelyn Dalfior

Tópicos

- Dataset
- Tarefas Aplicadas
 - Engenharia de Atributos
 - Classificação
 - Clustering
 - Regras de Associação
- Conclusões

Dataset

A Olist é uma empresa criada em 2007 como uma loja de shopping que depois se tornou um **marketplace** para a venda de produtos artesanais.

O conjunto de dados contém informações de **100 mil pedidos** realizados entre **2016 e 2018** em diversos marketplaces no Brasil. Suas features permitem visualizar um pedido sob múltiplas perspectivas: desde o status do pedido, preço, pagamento e desempenho do frete até a localização do cliente, atributos do produto e, por fim, avaliações escritas pelos clientes (9 arquivos no total).

Tarefas Aplicadas

Engenharia de Atributos: Construção de variáveis focadas na eficiência logística (frete, atraso e prazo):

- Tempo de aprovação em horas
- Tempo de entrega em dias
- Atraso de entrega em dias
- Tempo de transporte em dias
- Tempo de envio em dias
- Razão de frete
- Tempo estimado em dias
- Eficiência de entrega (dias de entrega sobre dias estimado de entrega)
- Diferença de dias de entrega (dia estimado x entrega real)

Tarefas Aplicadas - Engenharia de Atributos

Para melhorar a convergência do modelo linear e a qualidade das previsões, usamos padronização nos atributos numéricos e *class_weight="balanced"*.

Como *review_score* é ordinal e desbalanceado, uma alternativa é agrupar as notas em 2 classes:

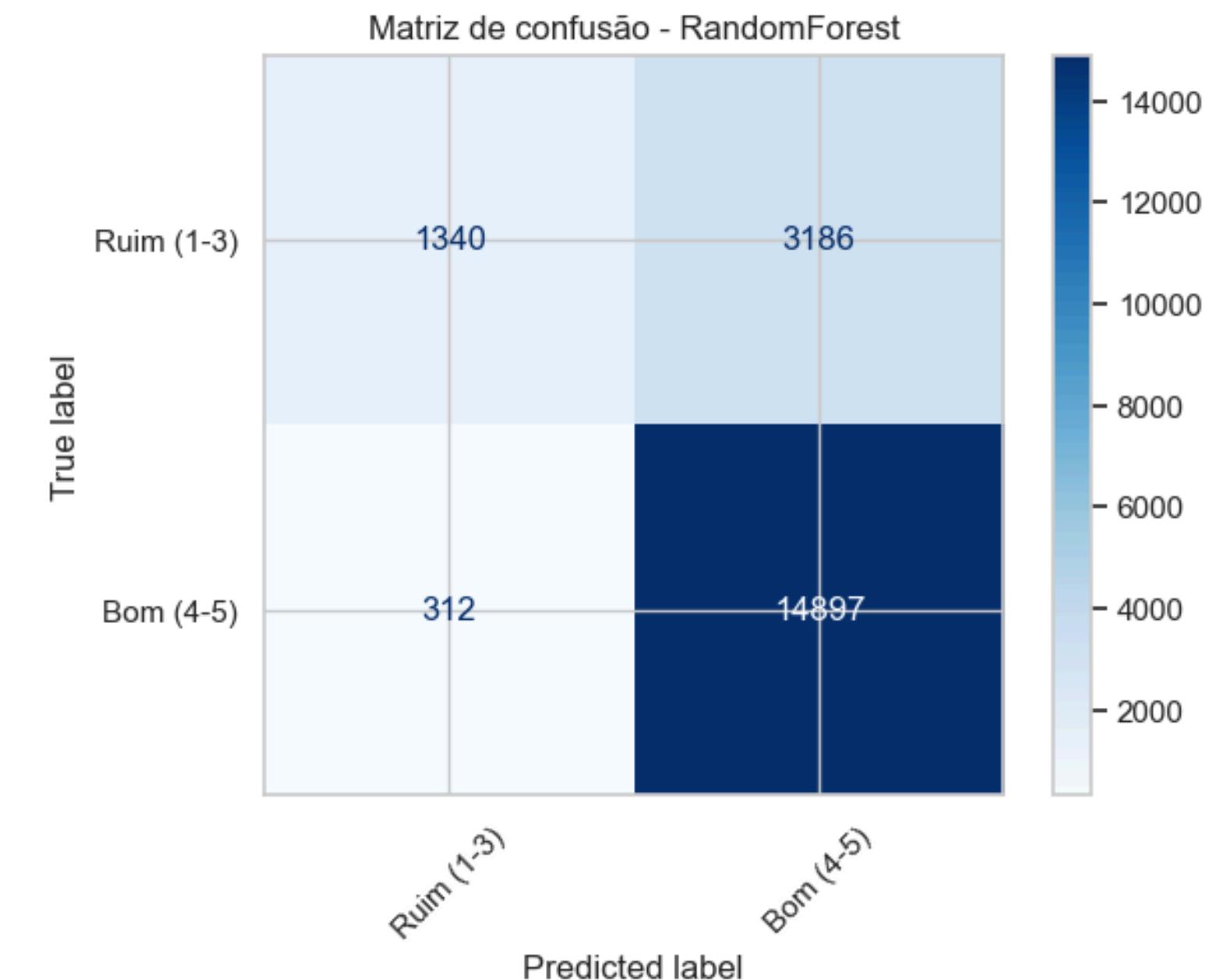
- 1, 2 e 3: negativo
- 4–5: positivo



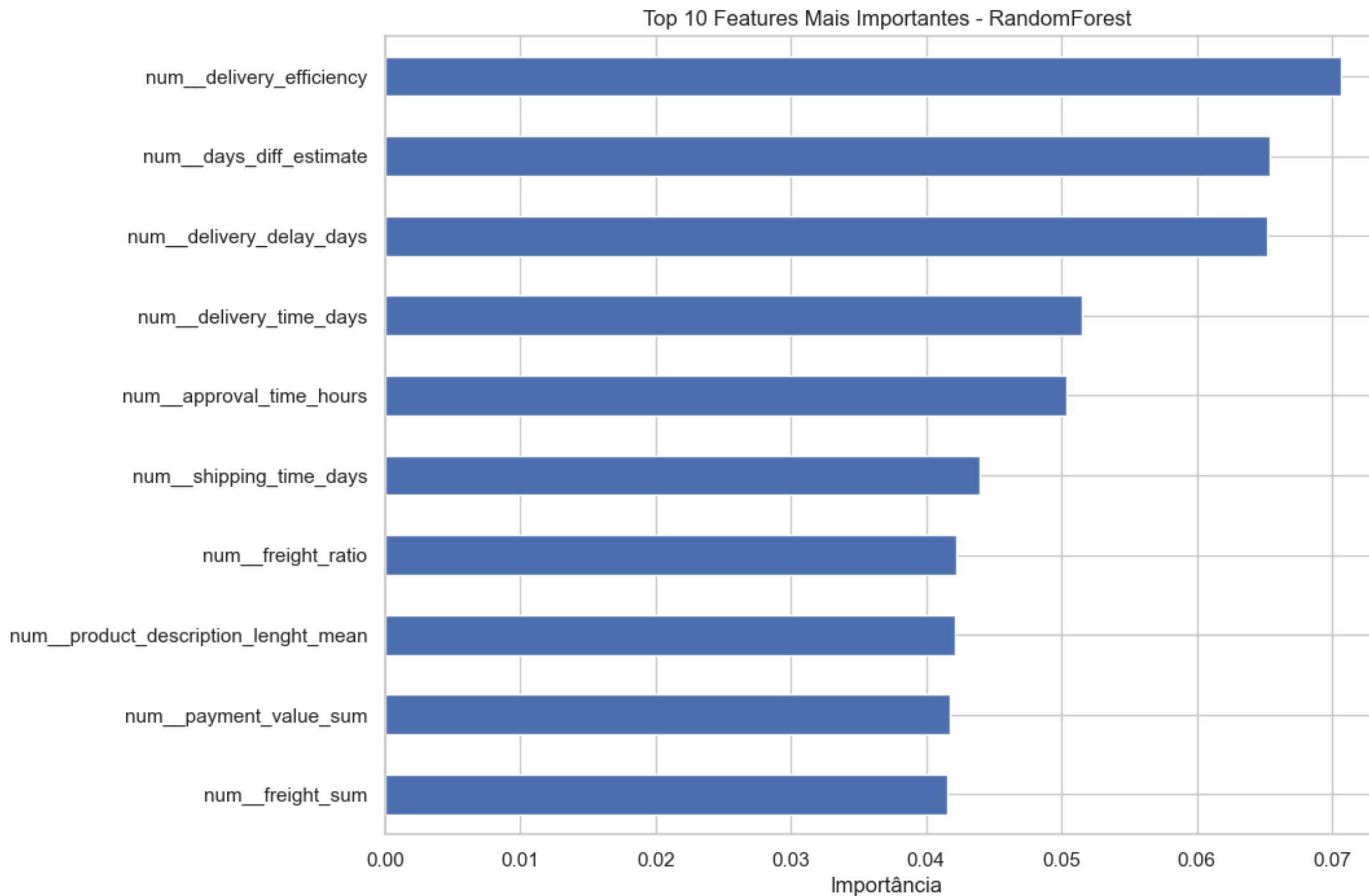
Tarefas Aplicadas

Classificação: Predição de risco de insatisfação do cliente utilizando algoritmos supervisionados, com foco nas métricas de precisão, revocação e matriz de confusão.

- **Random Forest:** acurácia de 82%, com 81% de precisão para a classe “Ruim”
 - Abordagem conservadora, identificando apenas 30% do total de problemas reais (recall).

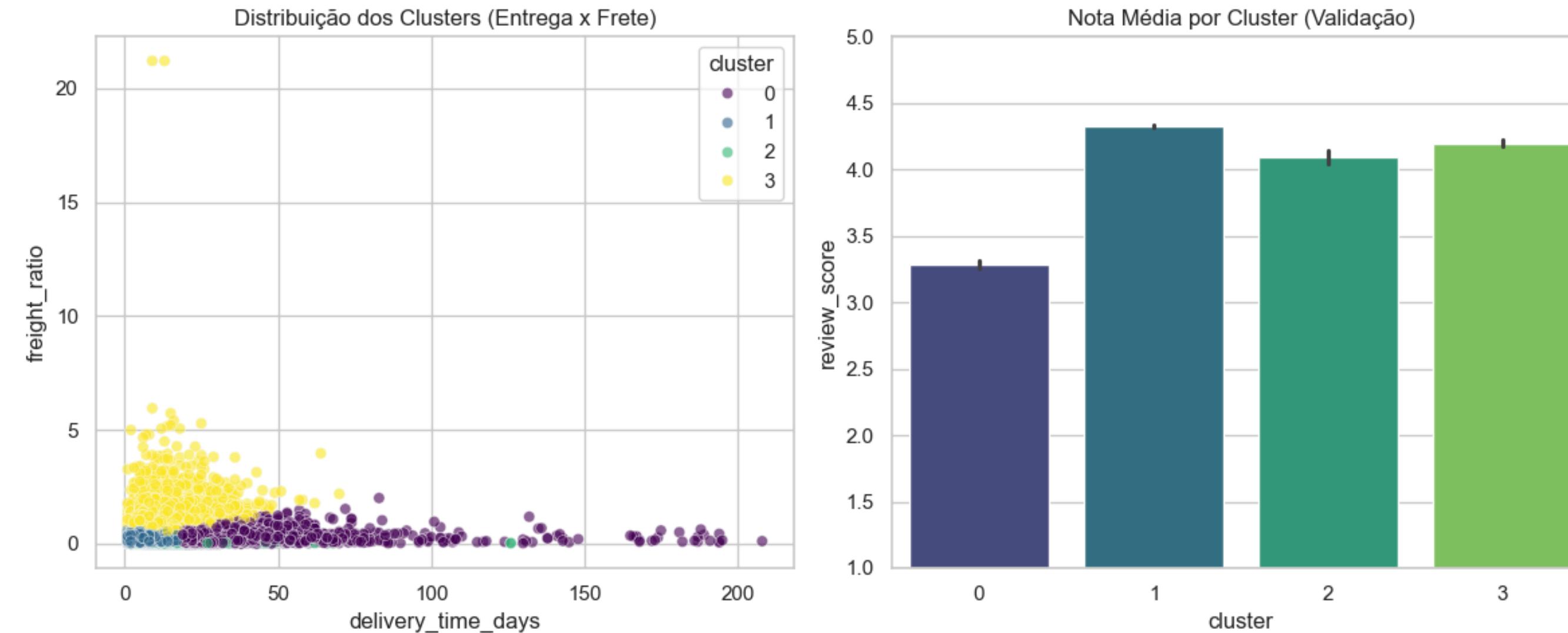


Tarefas Aplicadas - Classificação



Tarefas Aplicadas

Agrupamento (Clustering): Identificação de perfis logísticos não-supervisionados via K-Means, validando o impacto do tempo de entrega e do frete na experiência do usuário.



Tarefas Aplicadas - Clustering

- **Cluster 0: "Zona de Perigo Logístico"**

- **Características:** Demora quase um mês para chegar (28 dias). O frete é "ok" (27%), o preço é médio.
 - **Resultado:** Nota 3.28 (A pior de todas)

- **Cluster 1: "Experiência Ideal"**

- **Características:** Entrega "Jato" (8 dias). Preço acessível.
 - **Resultado:** Nota 4.32 (A melhor)

- **Cluster 2: "Ticket Alto / Frete Diluído"**

- **Características:** Produtos caros (Média R\$ 1.000,00). O frete representa muito pouco do valor (6%). Entrega média (12 dias)
 - **Resultado:** Nota 4.09 (Boa)

- **Cluster 3: "Produtos Baratos / Frete Caro"**

- **Características:** Produto muito barato (R\$ 29,00), mas o frete é quase o mesmo valor do produto (Ratio 0.94 ou 94%)
 - **Resultado:** Nota 4.19 (Surpreendentemente alta)

Tarefas Aplicadas

Regras de Associação: Descoberta de padrões de compra conjunta (Market Basket Analysis) utilizando o algoritmo Apriori para mapear oportunidades de *cross-selling*.

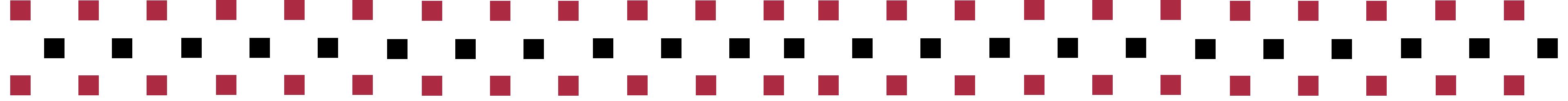
	antecedents	consequents	support	confidence	lift
14	frozenset({casa_conforto})	frozenset({cama_mesa_banho})	0.024412	0.739726	3.035759
15	frozenset({cama_mesa_banho})	frozenset({casa_conforto})	0.024412	0.100186	3.035759
21	frozenset({construcao_ferramentas_iluminacao})	frozenset({moveis_decoracao})	0.008137	0.600000	2.577087
20	frozenset({moveis_decoracao})	frozenset({construcao_ferramentas_iluminacao})	0.008137	0.034951	2.577087
2	frozenset({cool_stuff})	frozenset({bebes})	0.014014	0.196203	2.508671
3	frozenset({bebes})	frozenset({cool_stuff})	0.014014	0.179191	2.508671
0	frozenset({brinquedos})	frozenset({bebes})	0.012206	0.156069	1.995523
1	frozenset({bebes})	frozenset({brinquedos})	0.012206	0.156069	1.995523
22	frozenset({telefonia})	frozenset({informatica_acessorios})	0.009042	0.160000	1.734902
23	frozenset({informatica_acessorios})	frozenset({telefonia})	0.009042	0.098039	1.734902

Tarefas Aplicadas - Regras de Associação

- **Regra: casa_conforto → cama_mesa_banho**
 - Se um cliente compra algo de Casa Conforto (como uma almofada ou um puff), existe 74% de probabilidade (Confiança de 0.739) de ele também levar algo de "Cama, Mesa e Banho" (como um lençol ou toalha).
- **Regra construcao_ferramentas_iluminacao → moveis_decoracao**
 - Quem compra itens de iluminação (lustres, lâmpadas led), 60% também leva Móveis ou Decoração: cliente em um Momento de Compra (reforma?)
- **Regra telefonia → informatica_acessorios**
 - Lift acima de 1, pode indicar que um cliente ao comprar um celular, pode levar junto um cabo, carregador ou acessório de PC.
- **Bebês, Brinquedos e Cool Stuff**
 - Muitas regras que cruzem esses três itens, com lift alto (entre 1.99 e 2.50).

Conclusões

- O Random Forest provou ser uma ferramenta viável para **predição e ação proativa** (identificando pedidos de risco antes da avaliação ocorrer).
- A aplicação do algoritmo não-supervisionado **K-Means** corroborou, de forma independente, a hipótese de que a **logística** é o fator determinante para a **satisfação do cliente** no e-commerce brasileiro.
- A aplicação do **algoritmo Apriori** revelou que as cestas de compras na Olist não são puramente aleatórias. Foram identificados **padrões de comportamento** baseados no “momento de vida” do consumidor. Esses resultados podem ser diretamente aplicados na criação de “Kits de Produtos”, algoritmos de recomendação no carrinho de compras e campanhas de e-mail marketing direcionadas.



Obrigado!
Dúvidas?