

CIÊNCIA DE DADOS

MODELOS DE CLUSTERIZAÇÃO E CLASSIFICAÇÃO

Por Matheus Resende



01 Objetivo

Criar 2 modelos de machine learning aplicados a uma plataforma de delivery, aplicando conceitos de estatística, programação e análise e ciência de dados. O objetivo do 1º é realizar a segmentação de clientes para campanhas de marketing. Já o 2º utiliza como base a segmentação realizada pelo 1º para prever os clientes mais propensos a aceitar futuras campanhas.

02 Introdução

A segmentação de clientes é um dos pilares para o sucesso das plataformas de delivery, onde a variedade de perfis de consumidores e de estabelecimentos parceiros exige uma comunicação altamente direcionada.

Modelos de machine learning podem analisar o histórico de pedidos, frequência de compras e preferências culinárias para identificar padrões de comportamento e criar grupos de clientes com características semelhantes.

Essa abordagem permite que as campanhas de marketing sejam mais relevantes, oferecendo recomendações de restaurantes e promoções personalizadas que aumentam o engajamento e a taxa de conversão dos clientes.

03 EDA: Insights

Grupo 1 Clientes que aceitam mais campanhas		Grupo 2 Clientes que aceitam menos campanhas	
Qualitativo	Quantitativo	Qualitativo	Quantitativo
Renda média alta	Renda \geq 51.287,00	Renda média baixa	Renda $<$ 51.287,00
Faz mais compras	Compras \geq 12	Faz menos compras	Compras $<$ 12
Gasta mais em compras	Gastos em compras \geq 397,00	Gasta menos em compras	Gastos em compras $<$ 397,00
Não tem filhos	Filhos = 0	Tem filhos	Filhos \neq 0
Faz menos compras com desconto	Compras com desconto \leq 2	Faz mais compras com desconto	Compras com desconto $>$ 2
Visita menos o site	Visitas ao site \leq 6	Visita mais o site	Visitas ao site $>$ 6
Tem maior escolaridade	Escolaridade \neq Basic	Tem menor escolaridade	Escolaridade = Basic
Aceita mais campanhas	Aceitou campanhas = 1	Aceita menos campanhas	Aceitou campanhas = 0
Compra com mais frequência	Frequência de compras \leq 49	Compra com menos frequência	Frequência de compras $>$ 49
É cliente há mais tempo	Dias como cliente \geq 356	É cliente há menos tempo	Dias como cliente $<$ 356
Não possui um parceiro	Estado civil = Single	Possui um parceiro	Estado civil = Partner

04 EDA: Brainstorm

Cientes que aceitam campanhas

Renda média
alta

1

Faz mais
compras

2

Gasta mais em
compras

3



4

Tem maior
escolaridade

5

Faz compras
com mais
frequência

6

Aceita mais
campanhas

05 EDA: Projeção de Resultados*

Antes

- **15,1%:** Aceitariam futuras campanhas
- **84,9%:** Não aceitariam futuras campanhas



Depois

- **100,0%:** Aceitariam futuras campanhas
- **0,00%:** Não aceitariam futuras campanhas

* Essa projeção de resultados apresenta o potencial da adoção plena das soluções apontadas e considera que todas terão o resultado esperado. **No entanto, é preciso frisar que as condições aplicadas são para um grupo de clientes que está propenso a aceitar futuras campanhas e representa o potencial máximo desse estudo.** Dessa forma, é preciso entender com as demais áreas da empresa se há viabilidade na aplicação das ideias apresentadas ou se serão necessários mais estudos com novas soluções.

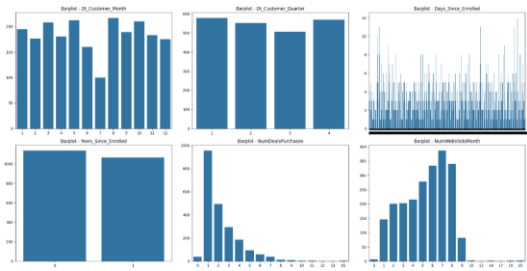
06 EDA: Potencial de Impacto

Cenário

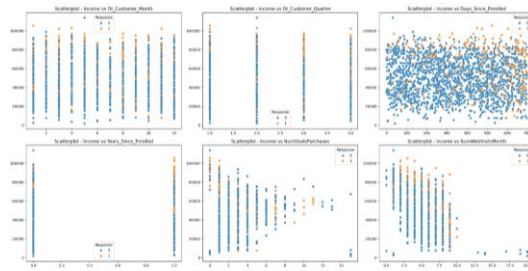
- **Receita:** R\$ 20 milhões
- **Conversão de campanhas:** + 10%
- **Faturamento adicional:** R\$ 2 milhões

07 EDA: Base Técnica

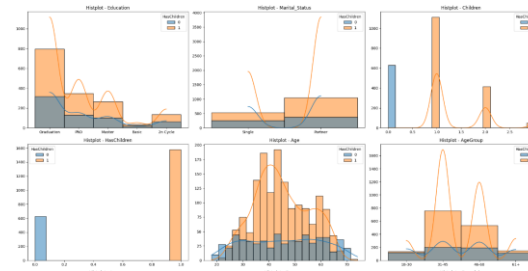
1: Gráfico de Contagem



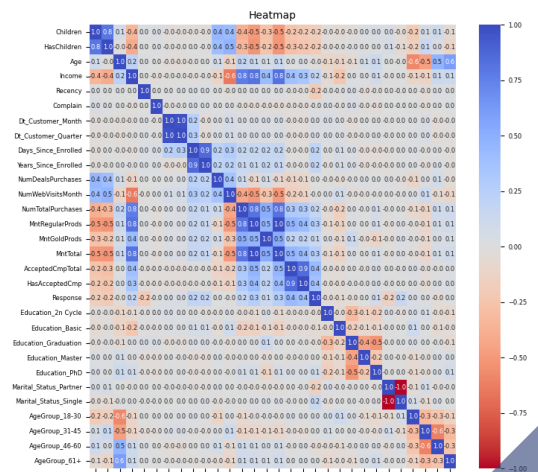
2: Gráfico de Dispersão



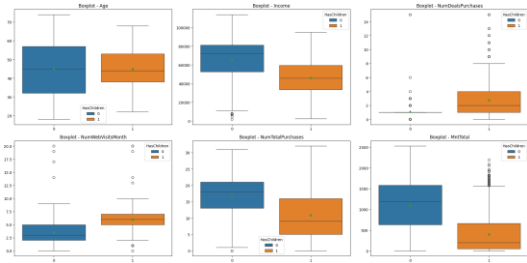
3: Histograma



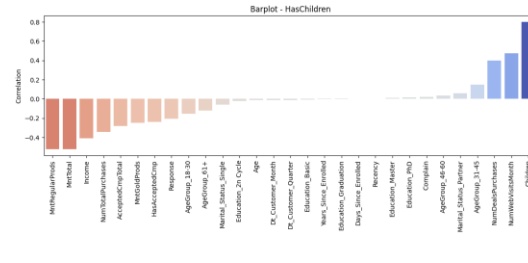
4: Mapa de Calor



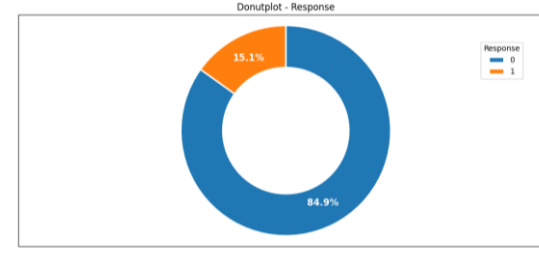
5: Diagrama de Caixa



6: Gráfico de Barra

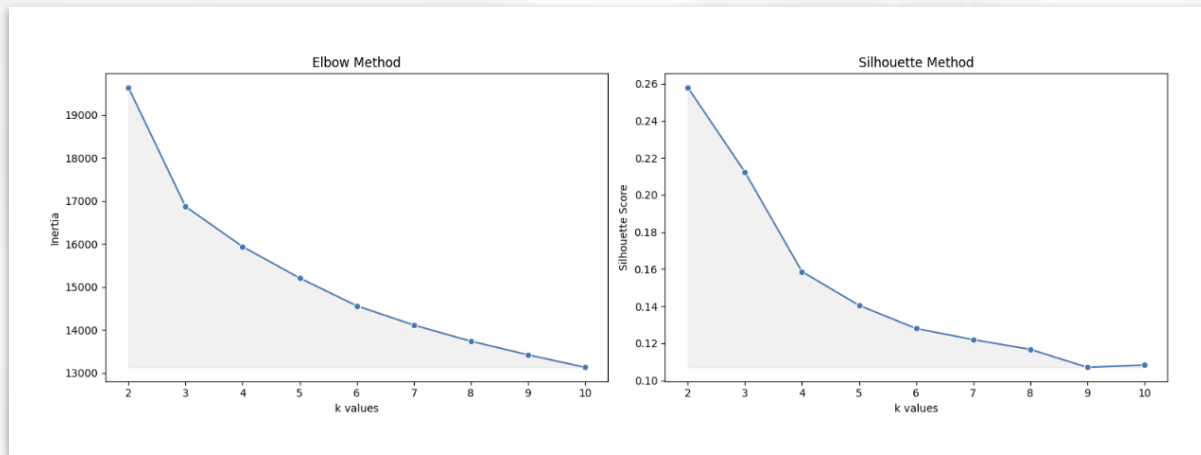


7: Gráfico de Rosca

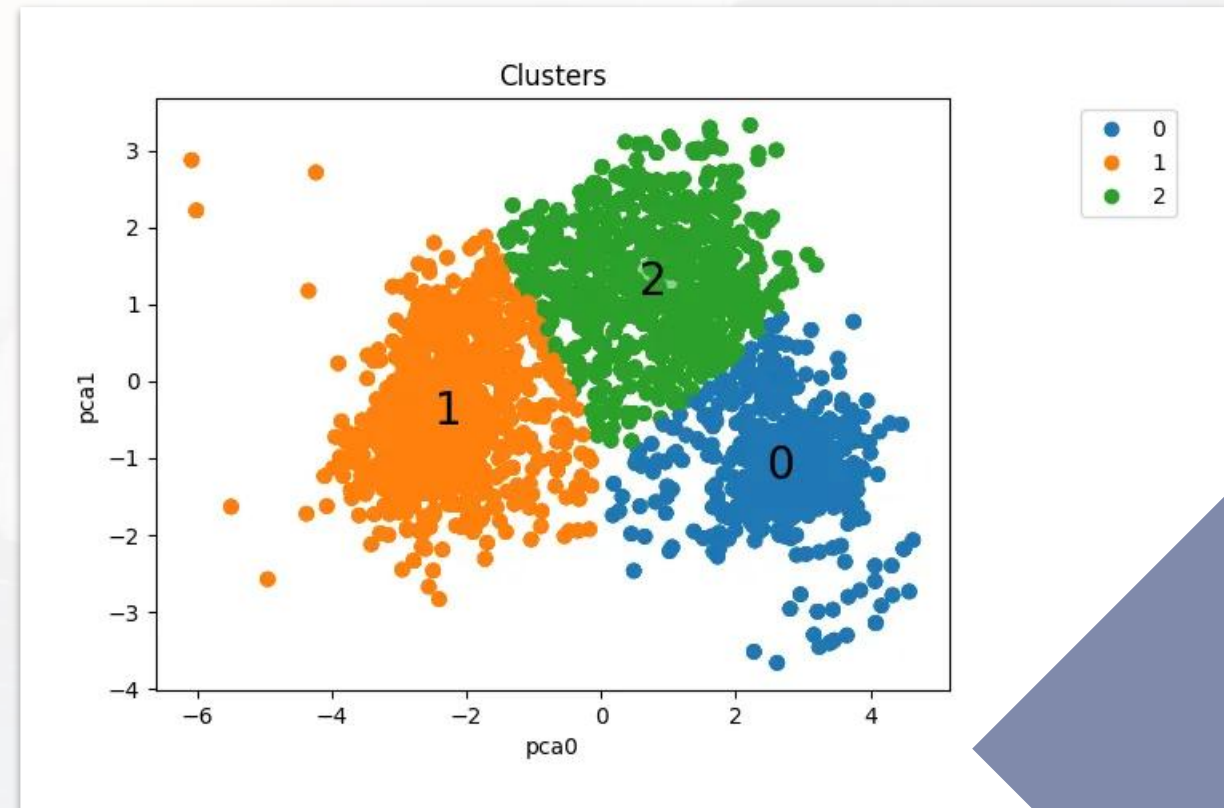


08 Modelo de Clusterização: K-means

1: Métodos do cotovelo e da silhueta

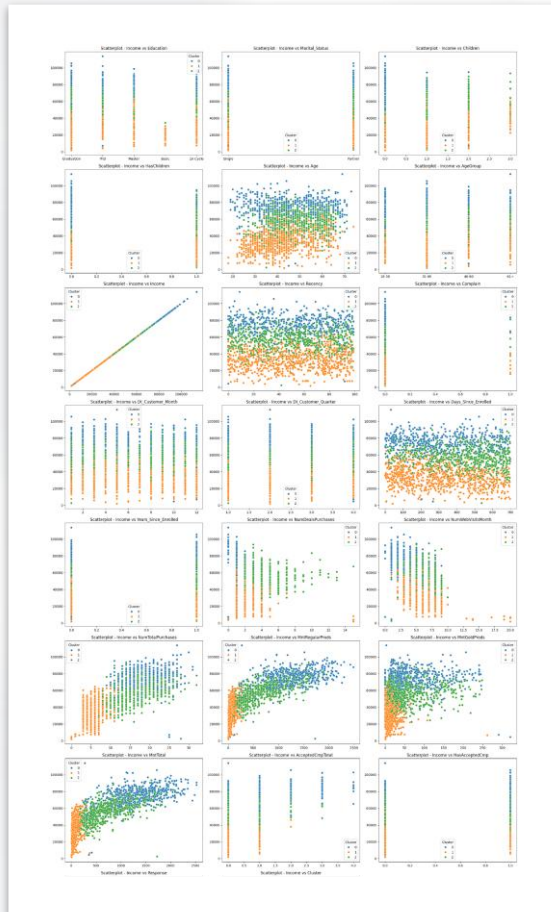


2: Modelo K-means

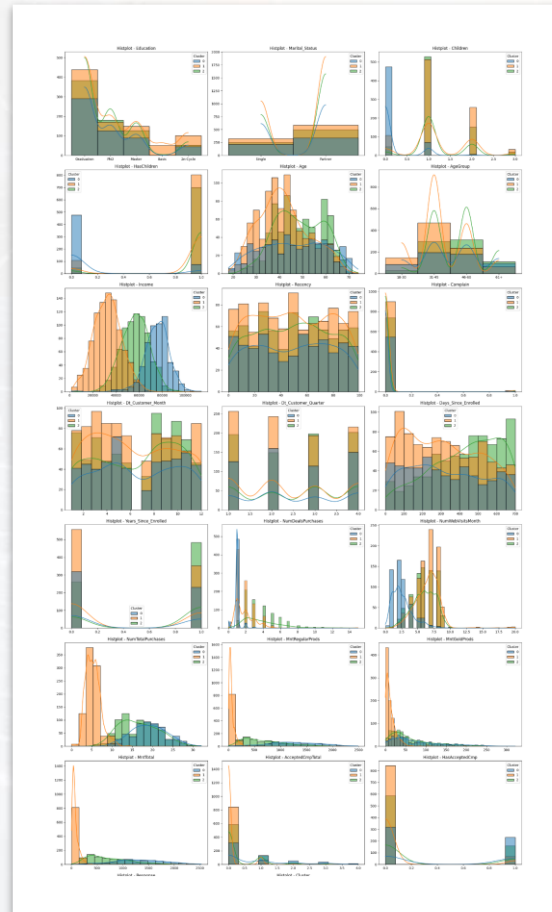


09 Modelo de Clusterização: Estatística

1: Gráficos de Dispersão



2: Histogramas



3: Diagramas de Caixa

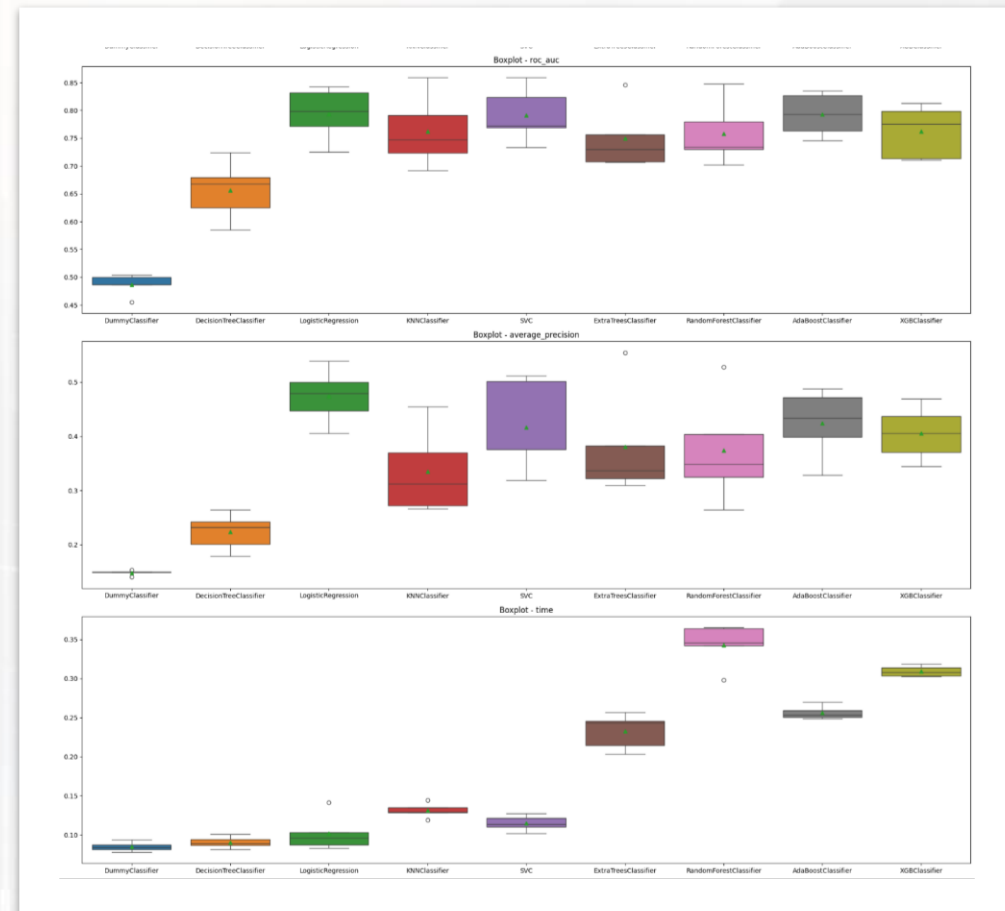
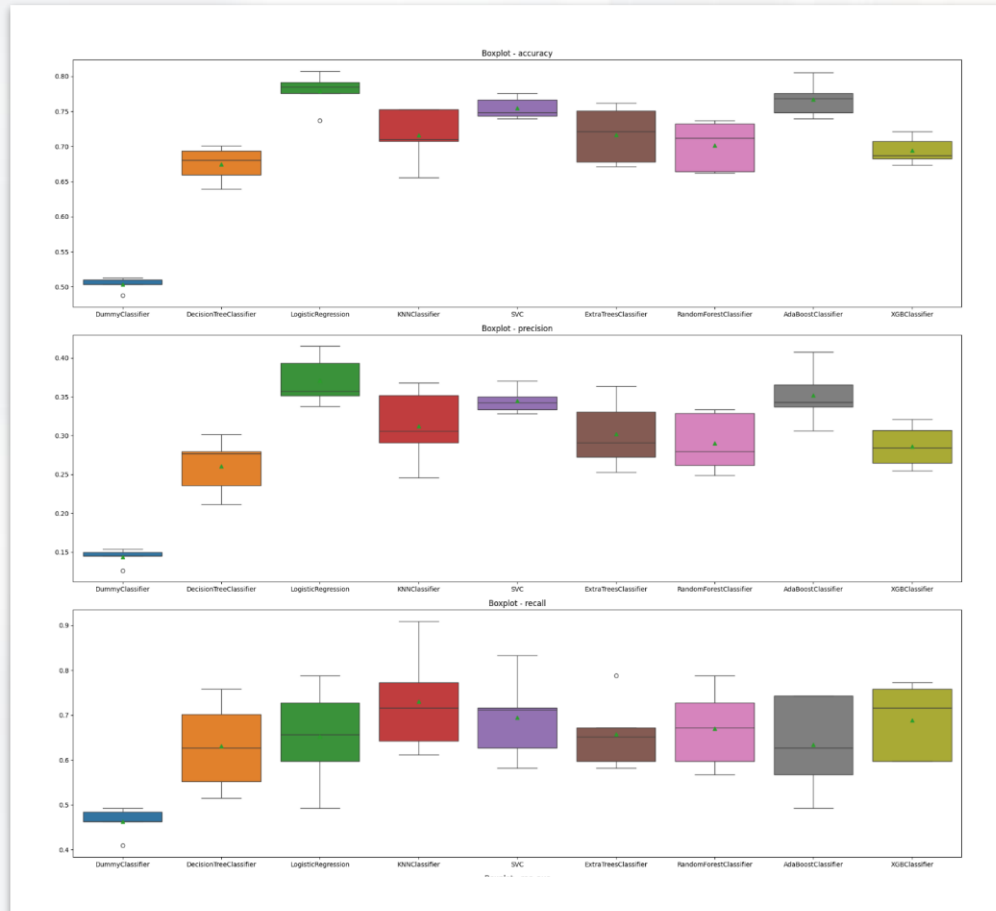


10 Modelo de Clusterização: Clusters

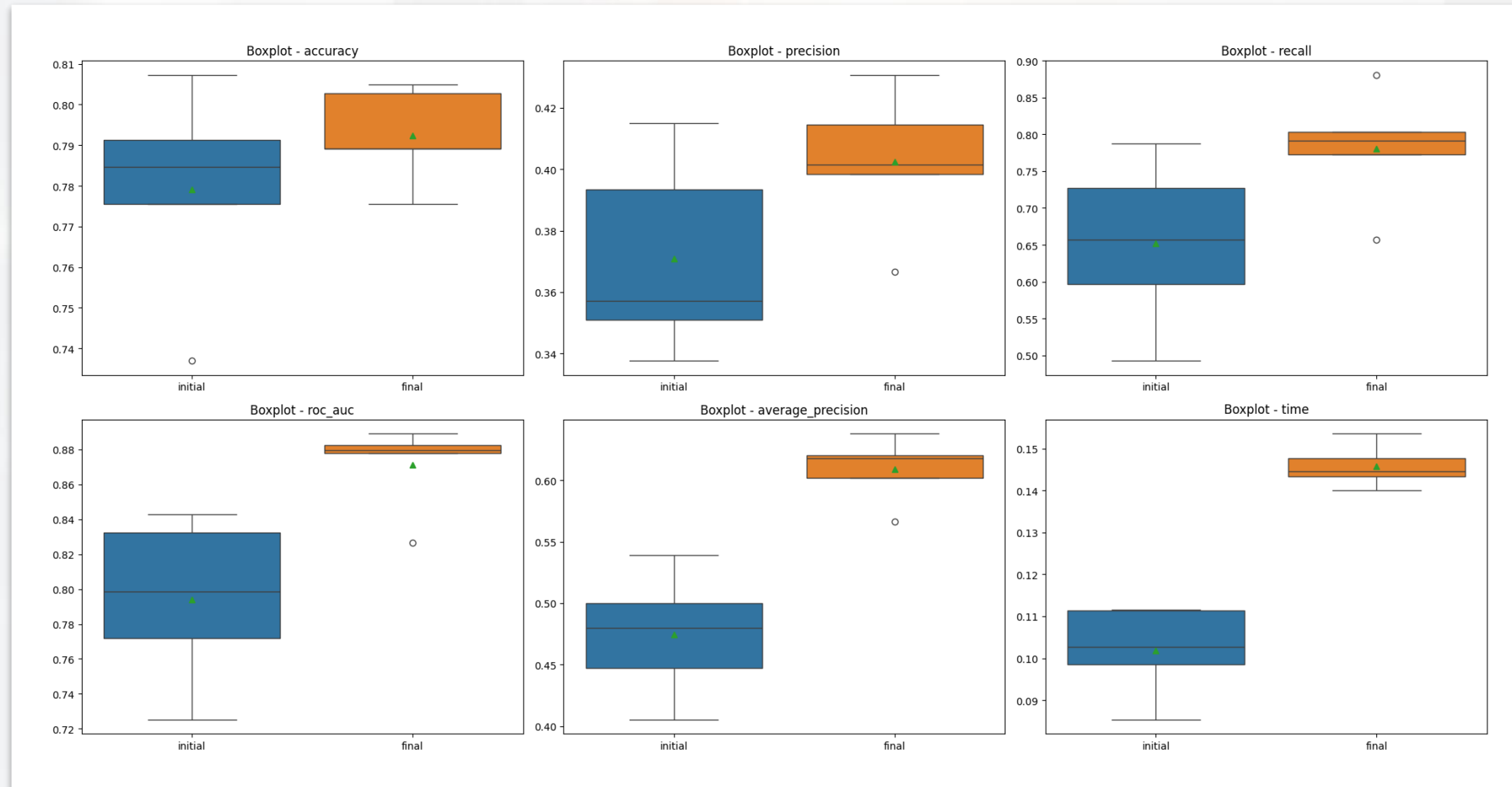
Cluster 0	Cluster 1	Cluster 2
Renda alta	Renda baixa	Renda média
Faz muitas compras	Faz poucas compras	Faz muitas compras
Gasta muito em compras	Gasta pouco em compras	Gasta moderadamente em compras
Provavelmente não tem filhos	Provavelmente tem filhos	Provavelmente tem filhos
Faz poucas compras com desconto	Faz algumas compras com desconto	Faz muitas compras com desconto
Visita pouco o site	Visita muito o site	Visita muito o site
Alta escolaridade	Único grupo que tem pessoas com escolaridade baixa	Alta escolaridade
Aceita campanhas facilmente	Difícilmente aceita campanhas	Pode aceitar campanhas
É cliente há menos tempo	É cliente há menos tempo	É cliente há mais tempo

11 Modelo de Classificação: Seleção

Modelo selecionado: Regressão Logística



12 Modelo de Classificação: Otimização



13 Overfitting

Boas práticas aplicadas

1. Seleção de features
2. Balanceamento de classes
3. Validação cruzada
4. Otimização de hiperparâmetros

```
3 require File.expand_path("../..//test_helper.rb", __FILE__)
4 # Prevent database truncation if the environment is production
5 abort("The Rails environment is running in production!") if Rails.env.production?
6 require 'spec_helper'
7 require 'rspec/rails'
8
9 require 'capybara/rspec'
10 require 'capybara/rails'
11
12 Capybara.javascript_driver = :webkit
13 Capybara.delete_all; Category.create
14 Matchers.configure do |config|
15   config.include_context :integration, :default
16   config.include_context :framework, :rails
17 end
```

14 Deploy em Produção

Modelo de Previsão

Prevendo a resposta da 6ª campanha

Escolaridade:

Graduation

Estado civil:

Single

Filhos:

0

Idade:

30

Renda:

88540,00

Gasto com produtos regulares:

750,00

Gasto com produtos gold:

150,00

Quantidade de campanhas aceitas:

2

Cluster:

☒ 0

☐ 1

☐ 2

FAZER PREVISÃO

Esse cliente DEVE aceitar a 6ª campanha.

15 Conclusão

Esse estudo evidencia como a combinação de modelos de machine learning de clusterização e classificação pode transformar a abordagem de marketing em plataformas de delivery, passando de apenas reativas para preditivas. Ao mapear perfis de clientes e antecipar comportamentos, a empresa está mais bem posicionada para desenvolver produtos, ofertas e estratégias que tenham sucesso com diferentes perfis de consumidores.



Obrigado