



# Problema de Negócio Hackathon Pod Academy



SQUAD 8

Álex Gabriel

Allan Basilio - PM

Jaqueline Jardim

Luís Alberto Costa

Marcos Souza

Pedro Melo - VA

Rafael Lima

Rafael Ribeiro

**Janeiro  
2025**

## 1. Arquitetura de Dados

Diagrama e principais descrições da arquitetura de dados

## 2. Modelo

Objetivo, Público-alvo, estratégias utilizadas e a escolha do modelo

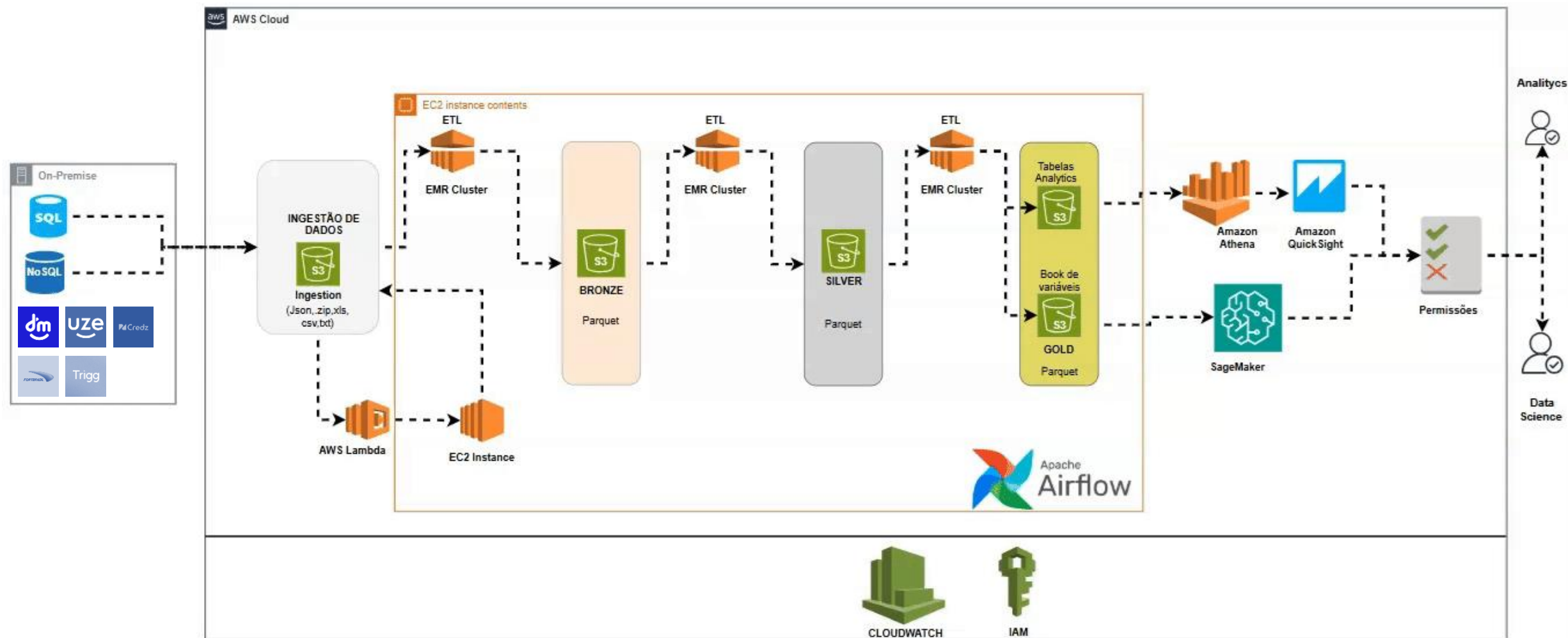
## 3. Resultados

Principais métricas, observações e monitoramento do modelo

## 4. Impacto do modelo para o negócio

Preservação de recurso e Aumento de Faturamento

# 1 Arquitetura de Dados



## 2 Modelo - Público Alvo

**Objetivo do modelo:** Classificar clientes elegíveis para aumento de limite, assim como aqueles com alto risco de inadimplência, atuando na redução de seus limites de crédito.

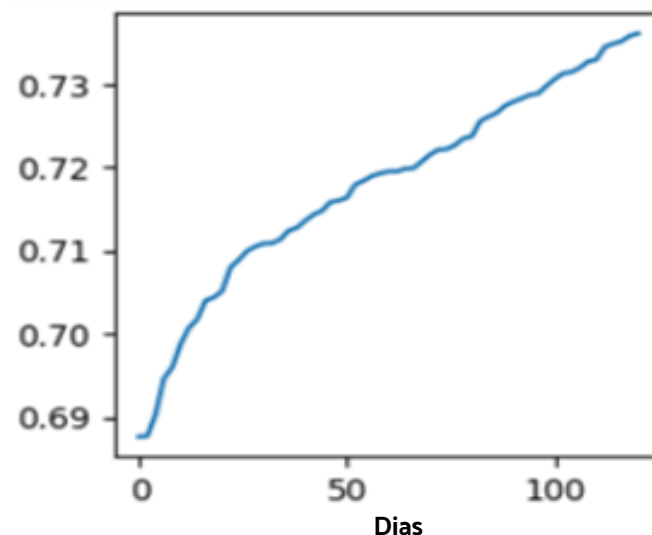
**Target:** over60mob6.

### Público alvo:

- Gasto nos últimos 6 meses;
- Atraso < 60 dias;
- Materialidade: Saldo > R\$ 50,00;
- Limite Não Nulo.



Evolução de dias de atraso em todas as carteiras



Faixa de Valores	Saldo Devedor Total	Porcentagem
R\$0-R\$25.00	R\$853,825.60	0.011591
R\$25.00-R\$50.00	R\$2,090,316.67	0.028378
R\$50.00-R\$100.00	R\$8,745,061.54	0.118723
R\$100.00-R\$200.00	R\$34,760,222.39	0.471903
R\$200.00-R\$300.00	R\$74,536,063.59	1.011898
R\$300.00-R\$400.00	R\$115,541,484.94	1.568585
R\$400.00-R\$500.00	R\$167,299,524.29	2.271250
R\$500.00-R\$1000.00	R\$1,056,248,850.37	14.339581
R\$1000.00-R\$2000.00	R\$1,698,375,184.26	23.057055
R\$2000.00-R\$5000.00	R\$2,612,517,976.22	35.467411
R\$5000.00+	R\$1,594,998,973.26	21.653625

### Determinação das safras:

Desenvolvimento : Jan/2023 → Dez/2023

Teste (OOS) : Jan/2023 → Dez/2023

VALIDAÇÃO (OOT) = Jan/2024 → Mar/2024



## 2 Modelo - Preparação da Tabela Analítica de Modelagem ( ABT)

**Objetivo da ABT:** Modelar o histórico do cliente nas empresas do Grupo DM de uma forma global e independente dos gastos por segmento.

### Tabelas Origem:

- ABT DM,
- ABT Credz,
- ABT Uze,
- ABT FortBrasil.

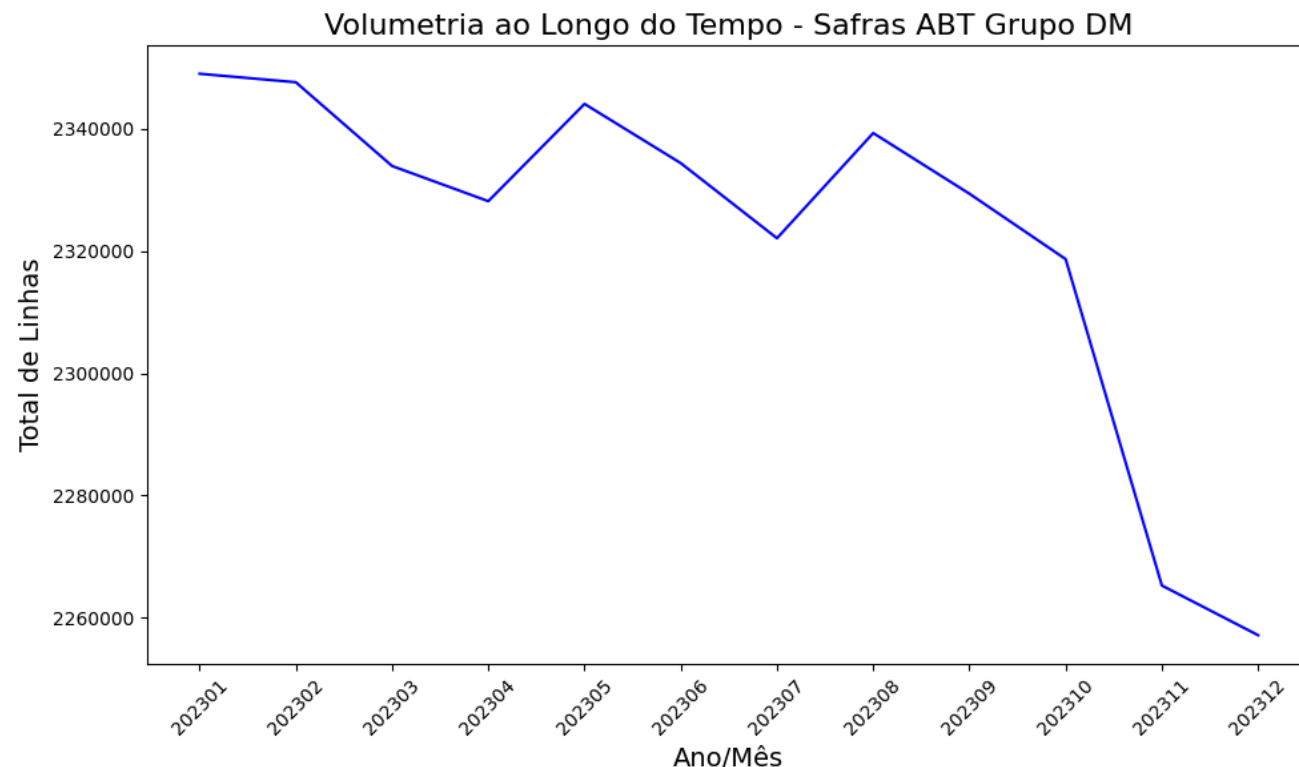
### Tabela Destino:

- ABT Grupo DM.

**Target:** Propagado das tabelas origem.

**Janela de Histórico:** 12 Meses

**Total de Features Criadas:** 64.



Criada a partir das ABTs da Credz, Uze, FortBrasil e DM. O objetivo desta tabela é modelar o histórico do cliente nas empresas do grupo DM de uma forma global e independente dos gastos por segmento. Foram desenvolvidas 64 features nessa ABT, modelando um único CPF por safra e com uma visão global das empresas do Grupo DM.

## 2 Modelo - Seleção das Variáveis

**Metodologia:** A seleção das variáveis para o modelo utilizou uma abordagem automatizada, utilizando **Random Forest**. Foi feita a remoção de variáveis com **multicolinearidade** e com variância baixa. A imputação nos dados foi feita usando **mediana** e a normalização feita usando **zscore**.

**Resultado:** Das 64, **18 variáveis foram selecionadas**. Incluindo indicadores de atraso, evolução do saldo devedor, variação do limite, razão de utilização do limite, quantidade de cartões, dias desde adesão do cartão.

Variável
dias_desde_adesao_no_grupo_dm
evolucao_saldo_1_2m_grupo_dm
evolucao_saldo_10_11m_grupo_dm
evolucao_saldo_11_12m_grupo_dm
evolucao_saldo_2_3m_grupo_dm
evolucao_saldo_3_4m_grupo_dm
evolucao_saldo_4_5m_grupo_dm
evolucao_saldo_5_6m_grupo_dm
evolucao_saldo_6_7m_grupo_dm
evolucao_saldo_7_8m_grupo_dm
evolucao_saldo_8_9m_grupo_dm
evolucao_saldo_9_10m_grupo_dm
max_dias_atraso_3m_grupo_dm
max_dias_atraso_no_grupo_dm
qntd_cartoes_total_no_grupo_dm
razao_utilizacao_limite_grupo_dm_12m
var_limite_6m_atras_no_grupo_dm
variacao_saldo_9m_grupo_dm

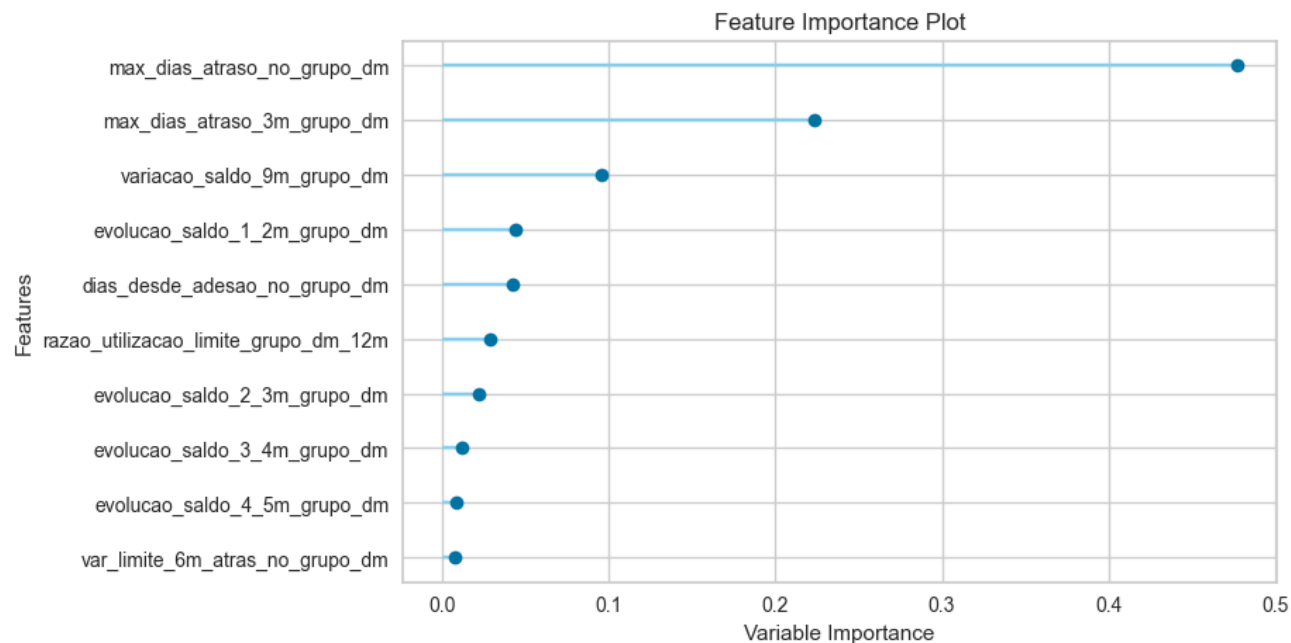
### 3 Resultados - Desenvolvimento

O **Gradient Boosting** foi o modelo campeão. Foi feita a otimização nesse modelo utilizando validação cruzada usando **5 folds**, resultando nas seguintes métricas.

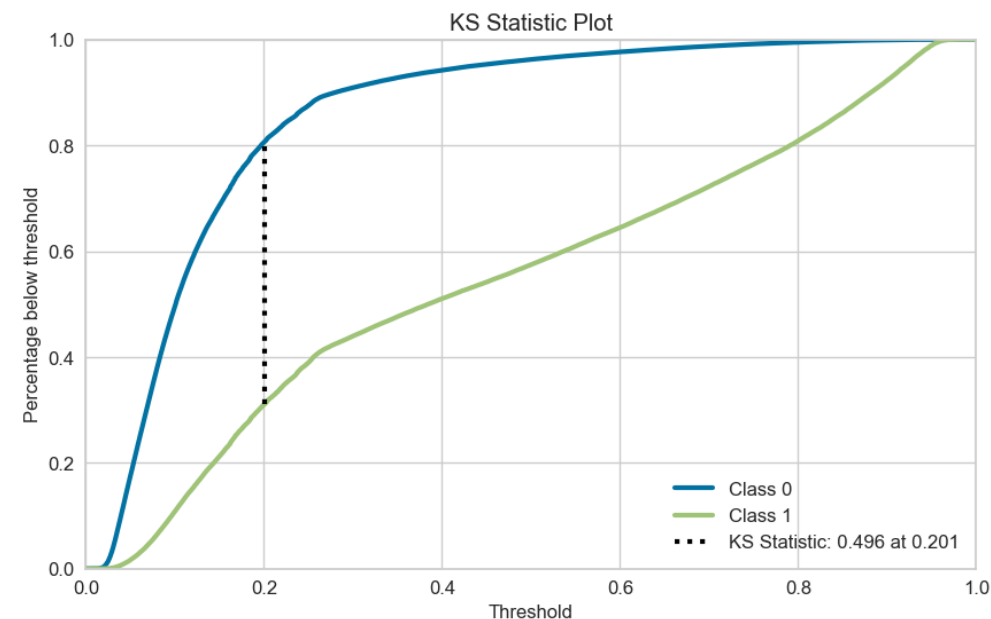
Modelo	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Gradient Boosting	0.8475	0.8177	0.4145	0.7278	0.5281	0.4452	0.4704

### Resultado ente Treino e Teste

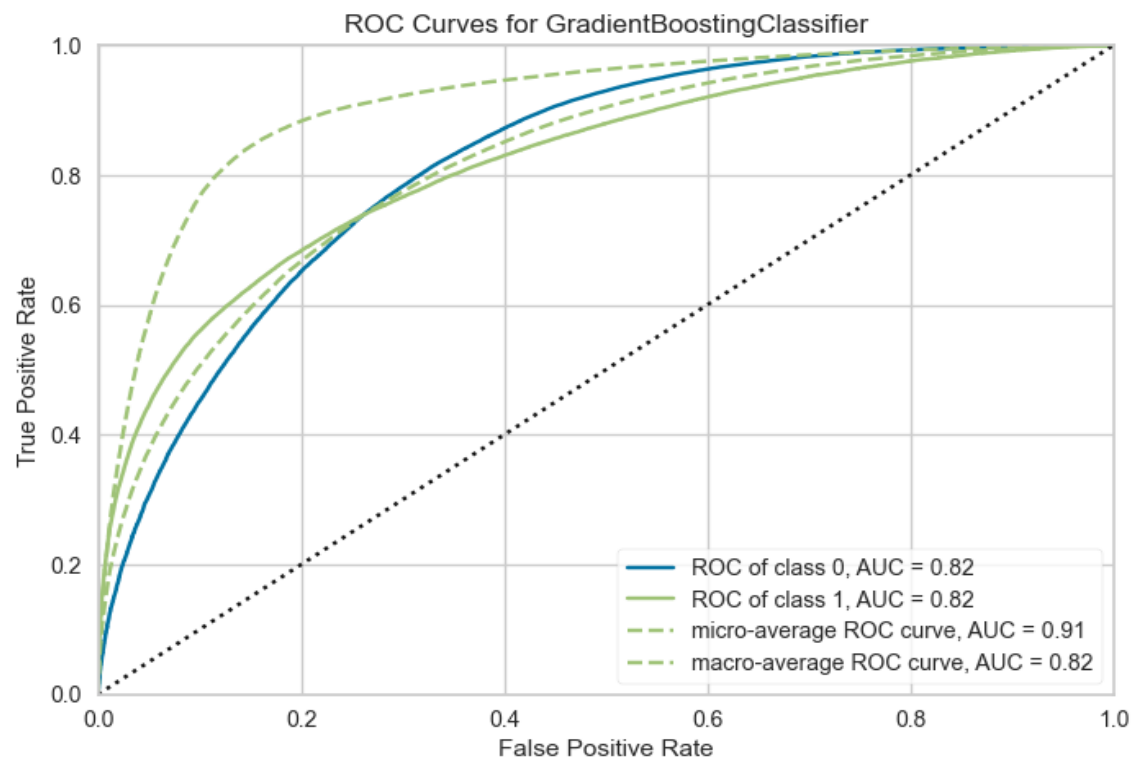
Modelo	AUC	Gini	KS
Gradient Boosting - Treino	0.83	0.65	0.49
Gradient Boosting - Teste	0.82	0.64	0.48



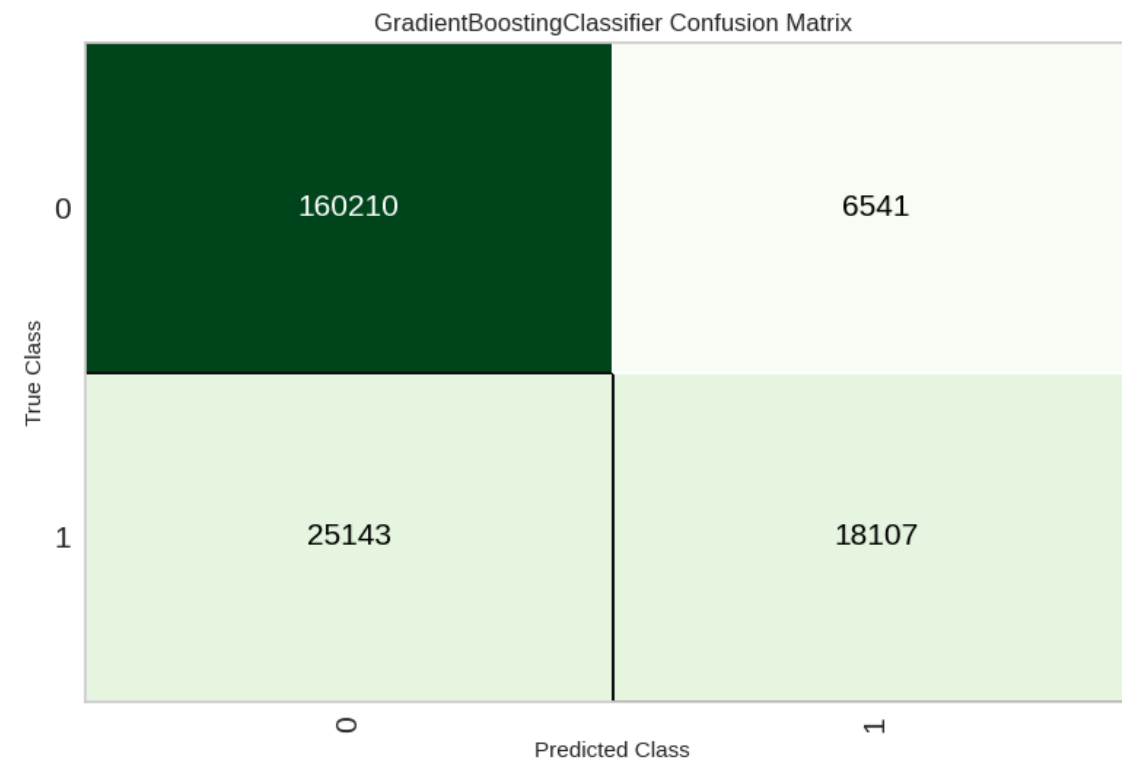
### Métrica: KS



## Métrica: AUC

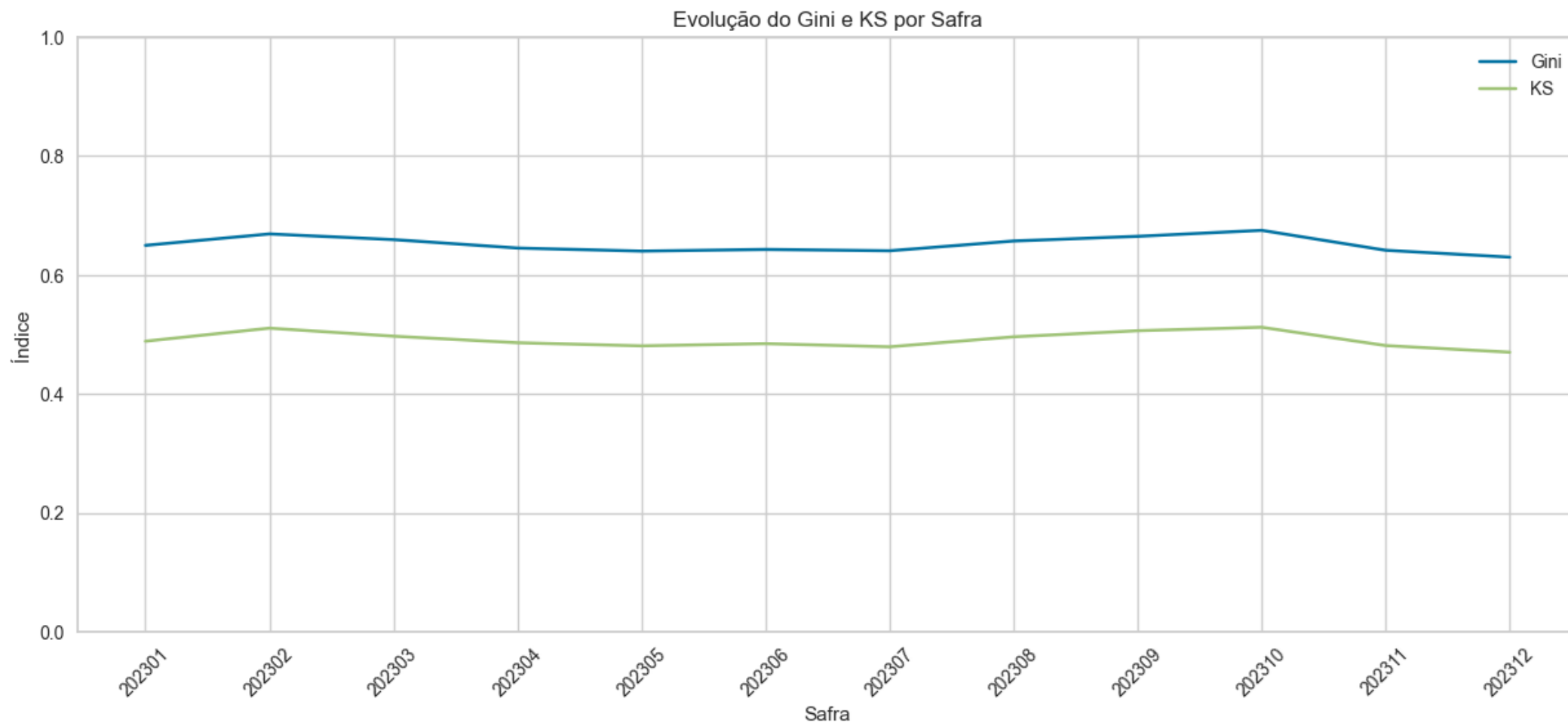


## Métrica: Matriz de confusão

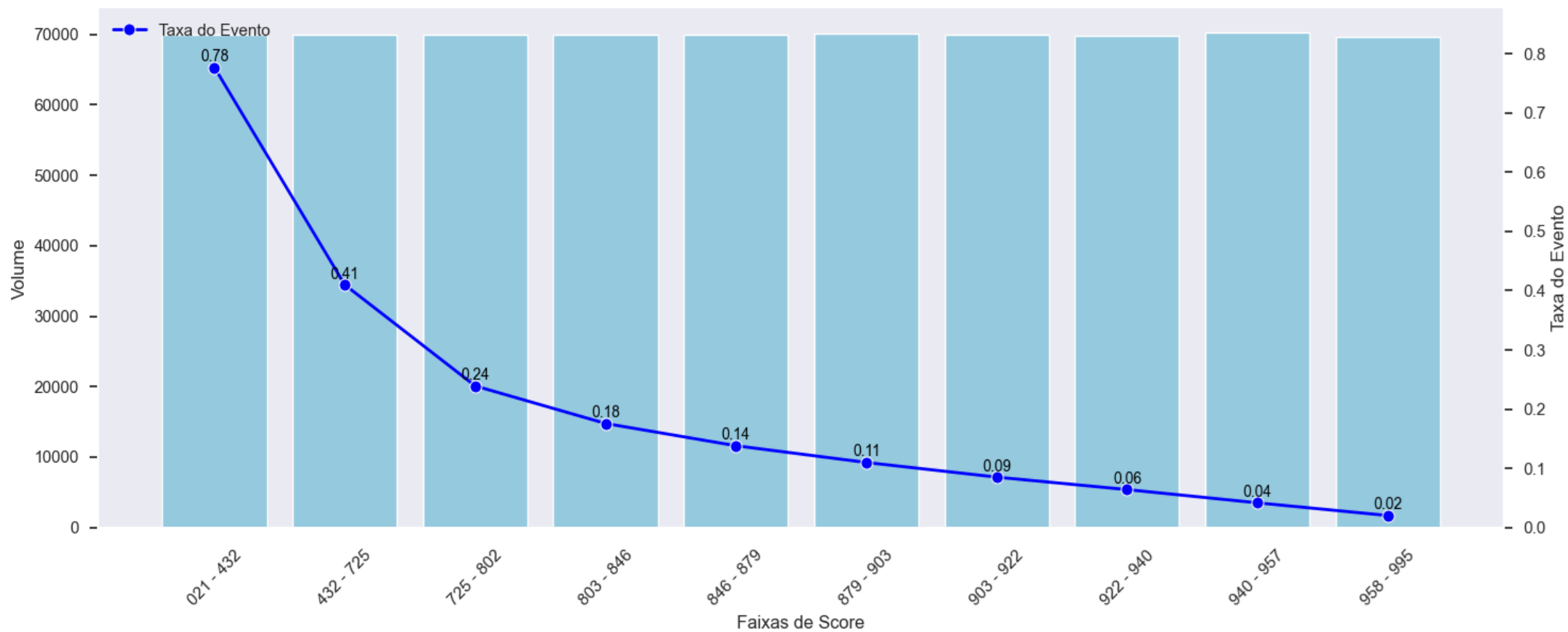




## A estabilidade do modelo ao longo das safras - Evolução do Gini e KS por Safra



## Taxa de Evento

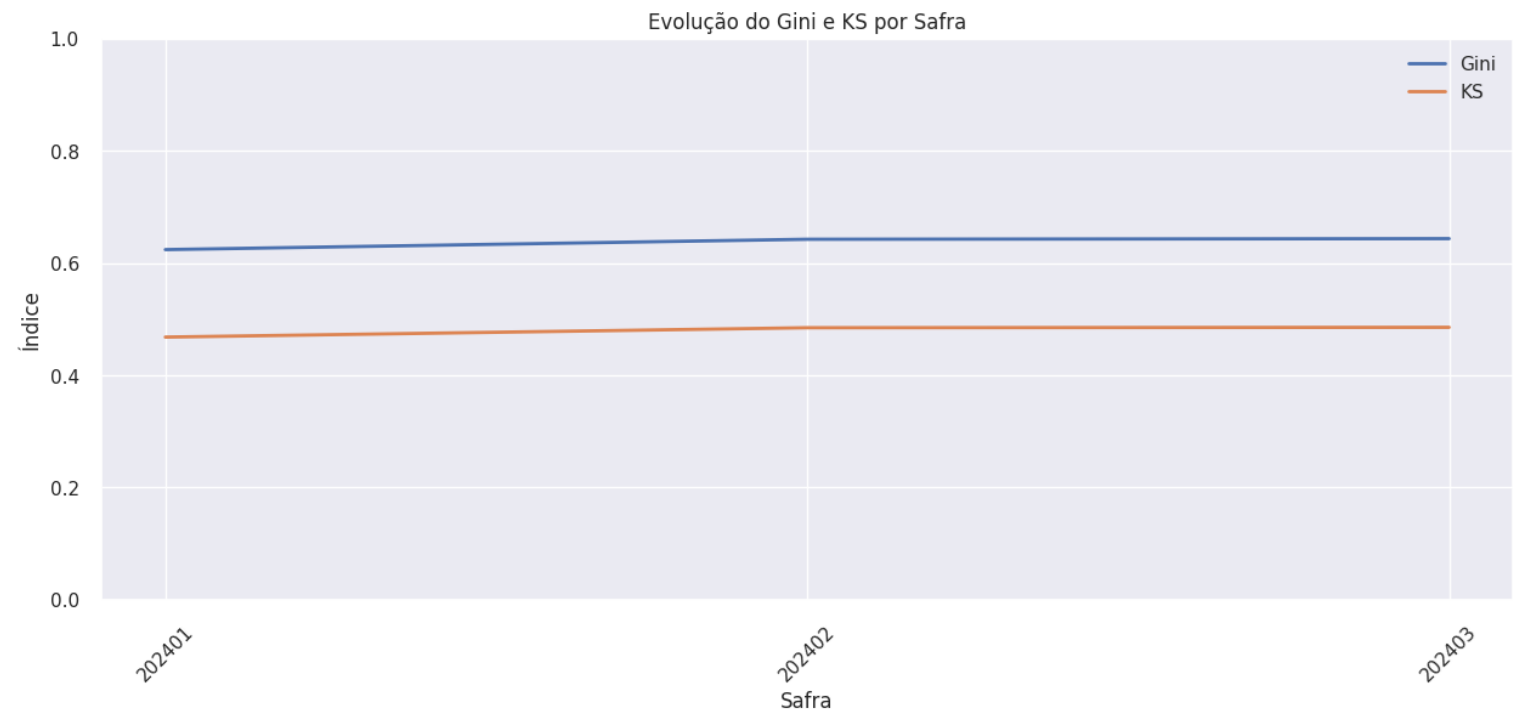
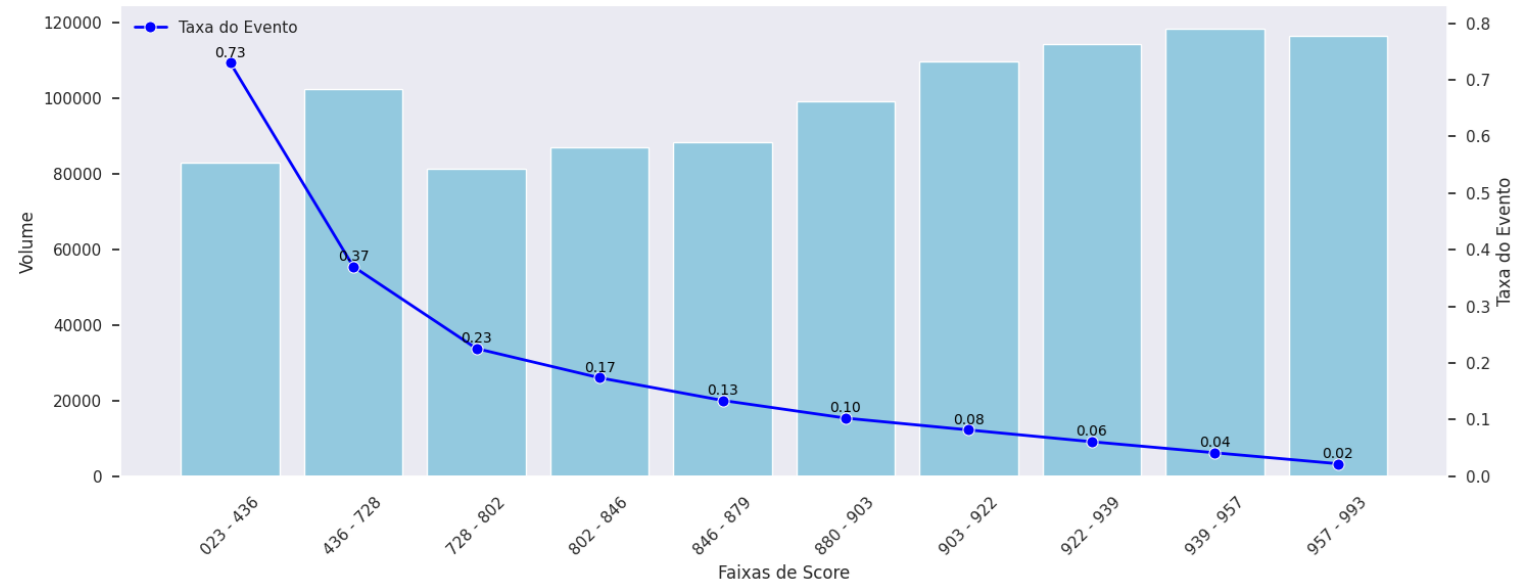


### 3 Resultados - Validação

#### Validação OOT

Modelo	AUC	Gini	KS
Gradient Boosting - Validação	0.82	0.64	0.48

A **estabilidade** observada nos índices entre o treinamento e a validação é uma evidência de que o modelo não sofreu overfit aos dados de treino, mas sim que ele tem a **capacidade de generalizar** bem para novos dados, um ponto crucial para a aplicação em produção.





## 3 Monitoramento do modelo

**Objetivo:** Garantir desempenho consistente (sustentação e melhorias) do modelo de behavior (over 60 mob 6) com detecção proativa de degradação.

### Estrutura AWS para Monitoramento:

- **SageMaker Model Monitor:** Detecta automaticamente drifts de dados e desempenho.
- **CloudWatch:** Centraliza métricas e gera alertas em tempo real.
- **S3:** Armazena logs, baseline do modelo e relatórios de métricas.
- **IAM:** Permissões otimizadas para segurança e controle de acesso.

### Benefícios do Plano de Monitoramento:

- Monitoramento automatizado;
- Re-treinamento baseado em evidências;
- Fácil escalabilidade;
- Integração com o ambiente atual da AWS.

## 1. Métricas de Desempenho

- **KS (Kolmogorov-Smirnov):** Para avaliar a separação entre classes (target e não-target).
- **GINI:** Para medir o poder de discriminação do modelo.
- **Distribuição de Previsões:** Comparar a distribuição atual das previsões (probabilidades) com a histórica (Data Drift).
- **Feature Importance Drift:** Monitorar mudanças significativas nas importâncias das variáveis.

## 2. Frequência das Verificações

- **Batch Mensal:** Verificações sincronizadas com a sumarização dos dados.
- **Automatização:** Configurar SageMaker Model Monitor para rodar inspeções no momento da execução batch.

## 3. Critérios de Re-treinamento

- **Métricas Abaixo do Limite:**
  - $KS < 0,4$
  - GINI ~ 15% em comparação com baseline.
- **Mudança na Distribuição dos Dados:**
  - Data Drift > 15% em variáveis-chave (comparação com baseline).
  - Feature Importance Drift > 10%.
- **Alertas Frequentes (ex.: 3 alertas consecutivos).**

## 4. Alertas de Degradação

- **AWS CloudWatch:**
  - Configurar alarmes baseados nos relatórios do **Model Monitor**.
  - **Categorias de alerta:**
    - **Alerta Crítico:** Métricas abaixo do limite ou detecção de data drift significativo.
    - **Alerta Informativo:** Tendências que indicam potencial degradação futura.
- **Integrações:** Notificações via Amazon SNS (e-mail, Slack ou Teams).



## 4 Impacto do modelo para o negócio

R\$ 9,4 Milhões poderiam ser **preservados** ao aplicar a regra de negócio a partir dos resultados do modelo.

Utilizando o modelo na safra de Janeiro 2024

Obtivemos 69,3% de acurácia (Predict label)

Selecionando as primeiras 5 faixas de score 024 a 509, temos 100.887 clientes:

total_limite_acumulado	total_saldo_devedor	media_limite_acumulado	media_saldo_devedor
118,029,127.440	188,895,483.000	1,169.914	1,872.347

Se aplicarmos a **redução de limite em 30%** sobre esse total de clientes:

total_limite_acumulado_reduzido	valor_reduzido_30
82,620,389.21	35,408,738.23

Ao filtrar 20.653 clientes com o **saldo devedor menor que o limite** em jan/24, seria possível **retirar R\$ 9.394.579.36** de recurso da zona de maior risco.

total_cpfs	total_limite	total_saldo_devedor	total_recurso_preservado
20653	32,064,436.42	22,669,857.06	9,394,579.36

fx_score_cat	qtde_cpfs_distintos
024 - 121	14518
121 - 218	19799
218 - 315	19980
315 - 412	22749
412 - 509	23841
509 - 606	29618
606 - 703	39103
703 - 800	93335
800 - 897	258998
897 - 994.9	478059



## 4 Impacto do modelo para o negócio

R\$ **63,4** Milhões de **aumento no limite** ao aplicar a regra de negócio a partir dos resultados do modelo.

Utilizando o modelo na safra de janeiro 2024

Obtivemos **91,3%** de acurácia (Predict label)

fx_score_cat		qtde_cpfs_distintos
024 - 121		14518
121 - 218		19799
218 - 315		19980
315 - 412		22749
412 - 509		23841
509 - 606		29618
606 - 703		39103
703 - 800		93335
800 - 897		258998
897 - 994.9		478059

Nas duas últimas faixas de **score 800 a 994.9**, temos **737.057** clientes.

Se aplicarmos o aumento de limite de **30%** sobre esse total de clientes:

total_limite_com_aumento	valor_aumento_30
1,589,495,057.54	366,806,551.74

Para não conceder esse aumento aos cliente que não fariam uso, foi aplicado uma **seleção daqueles que tinham saldo devedor igual ou maior que seus limites**:

total_cpfs	total_limite	total_saldo_devedor
199131	211,306,991.12	343,290,438.88

\* 62,5% de diferença entre saldo devedor e total de limite.

Aplicando o aumento somente para os **100.131** clientes:

total_cpfs	total_limite	total_limite_com_aumento	valor_aumento_30
199131	211,306,992.00	274,699,089.60	63,392,097.60





dm | Obrigado! ✨

[www.dmcard.com.br](http://www.dmcard.com.br)



0800 702 5004



@dmcard



@dmcard



@dmcardoficial