

credbank

CREDIT SCORE

Gerência de Ciência de Dados



Sobre nós

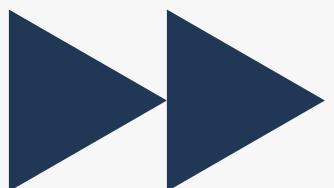
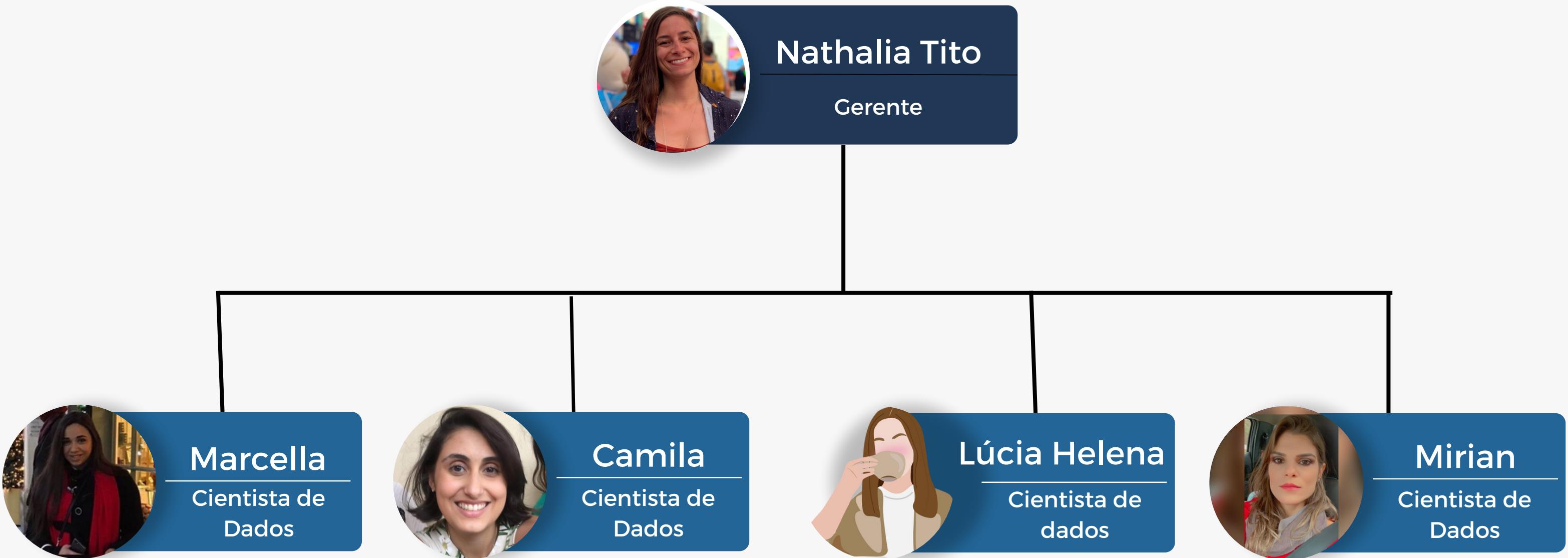
Somos uma plataforma digital na qual os nossos profissionais de negócios e clientes têm à disposição um ecossistema de produtos financeiros e conveniências.

A nova área de ciência de dados foi criada para que, através do uso de diversas tecnologias, modelos e metodologias seja possível **capturar, armazenar e processar informações**, gerando valor ao nosso negócio.



credbank

Organização da área



Problema



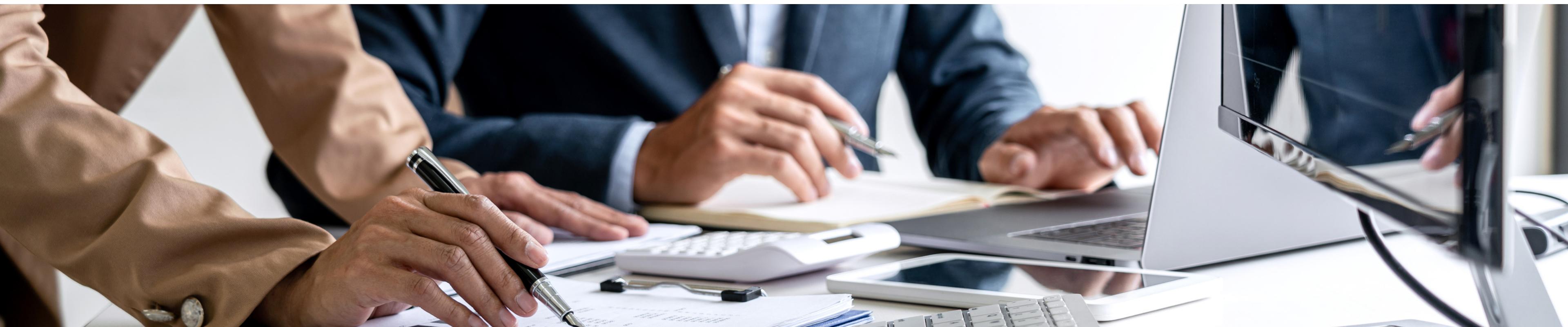
Inadimplência 20% acima
da meta



Alto PCLD



84% do valor da dívida com
aging acima de 180 dias



Causas

1

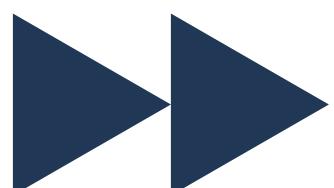
O crédito é liberado apenas de acordo com o salário do cliente

2

Não existe análise de importância de features

3

Critérios definidos por feeling



Objetivos

Construção de um modelo de aprendizado de máquina para classificar a pontuação de crédito reduzindo os esforços manuais.



Resposta para as seguintes perguntas:

O que faz alguém ter um score maior ou menor, a quantidade de empréstimos é relevante? O saldo mensal?

Quais dados temos disponíveis?

Temos uma base com informações de **100.000** clientes devedores e **24** Features relacionadas à eles, como:

Idade

Quantidade de cartões de crédito e contas bancárias

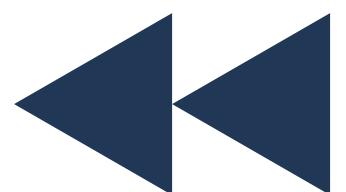
Renda Mensal e anual

Média de atraso

Valor em atraso

Quantidade e tipos de empréstimo

Idade do histórico de crédito



Precisávamos usar todas essas informações?

NÃO NECESSÁRIAMENTE!



Com o modelo, é possível entender quais são as variáveis que são realmente importantes para a definição do *credit score* e são essas as variáveis (*features*) que usamos para construirmos o nosso modelo !

Personas

Poor



NOME

HARLOW

Standard



WESLEY

Good



RAYNE

PROFISSÃO

Estudante

Engenheiro

Jornalista

IDADE

25 anos

32 anos

35 anos

TIPO DE EMPRÉSTIMO

Crédito Estudantil

Construtora

Casa Própria

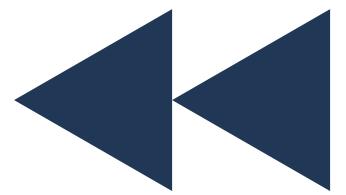
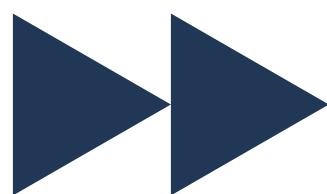
SALÁRIO MENSAL

R\$ 1.700,00

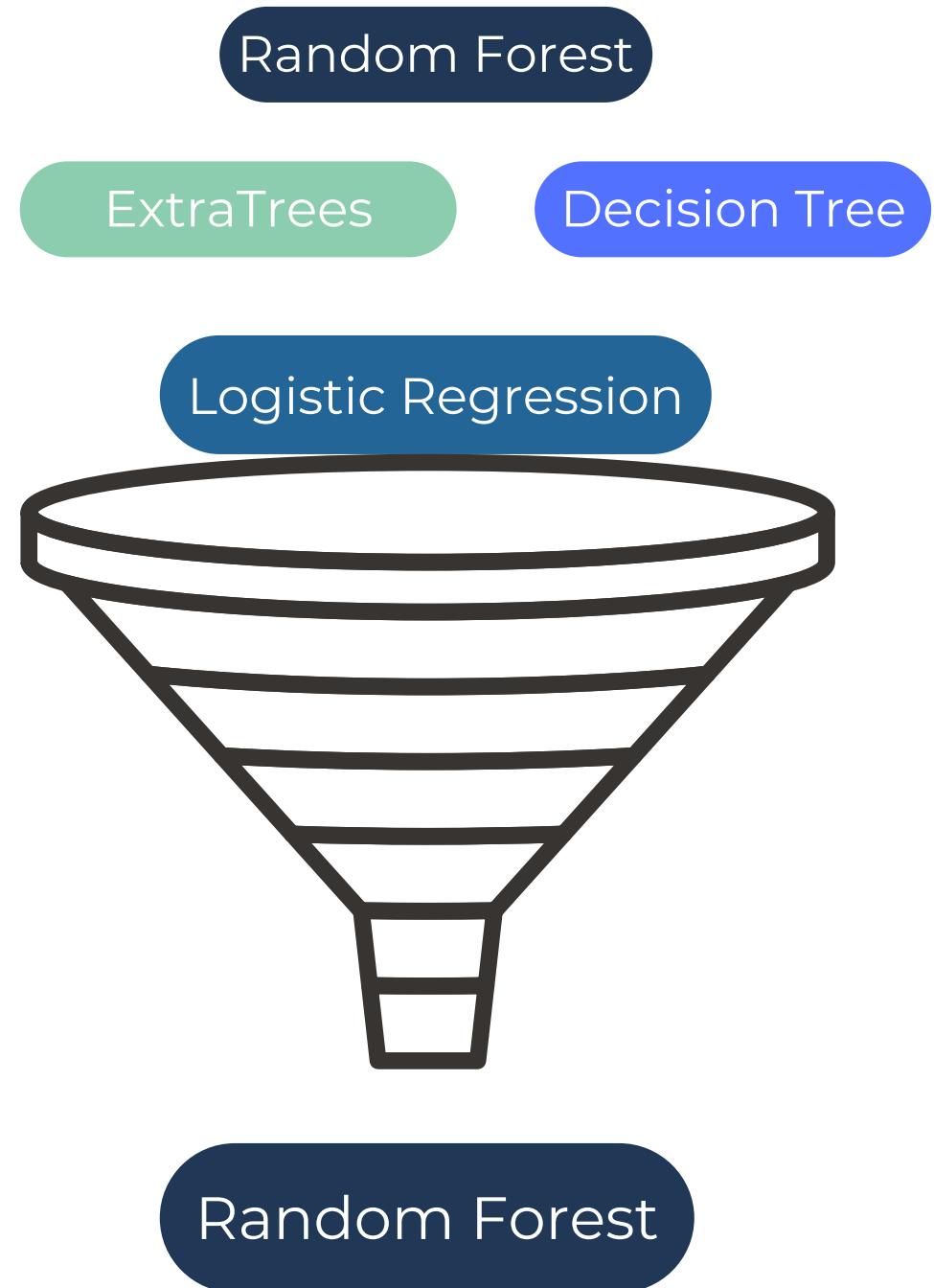
R\$ 6.000,00

R\$ 4.750,00

Etapas do processo



Modelo escolhido



Resultados

Modelo LogisticRegression:

Precision:69.53%

Classification Report:

	precision	recall	f1-score	support
0.0	0.47	0.83	0.60	2612
1.0	0.64	0.66	0.65	4132
2.0	0.80	0.58	0.67	7728
accuracy			0.65	14472
macro avg	0.64	0.69	0.64	14472
weighted avg	0.70	0.65	0.65	14472

Modelo ExtraTreesClassifier:

Precision:79.80%

Classification Report:

	precision	recall	f1-score	support	
0.0	0.0	0.71	0.81	0.76	2612
1.0	1.0	0.77	0.83	0.80	4132
2.0	2.0	0.84	0.77	0.80	7728
accuracy				0.79	14472
macro avg	0.77	0.80	0.79	0.79	14472
weighted avg	0.80	0.79	0.79	0.79	14472

Modelo RandomForest:

Precision:80.11%

Classification Report:

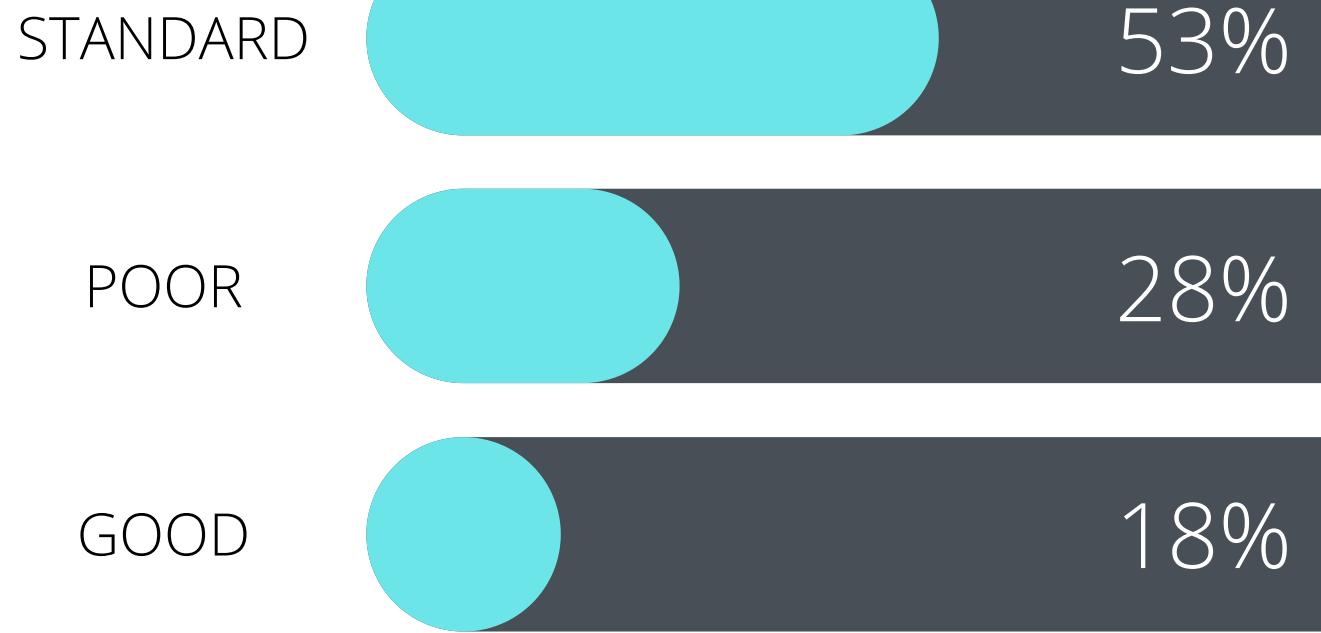
	precision	recall	f1-score	support
0.0	0.72	0.80	0.76	2612
1.0	0.77	0.84	0.80	4132
2.0	0.84	0.78	0.81	7728
accuracy			0.80	14472
macro avg	0.78	0.80	0.79	14472
weighted avg	0.80	0.80	0.80	14472



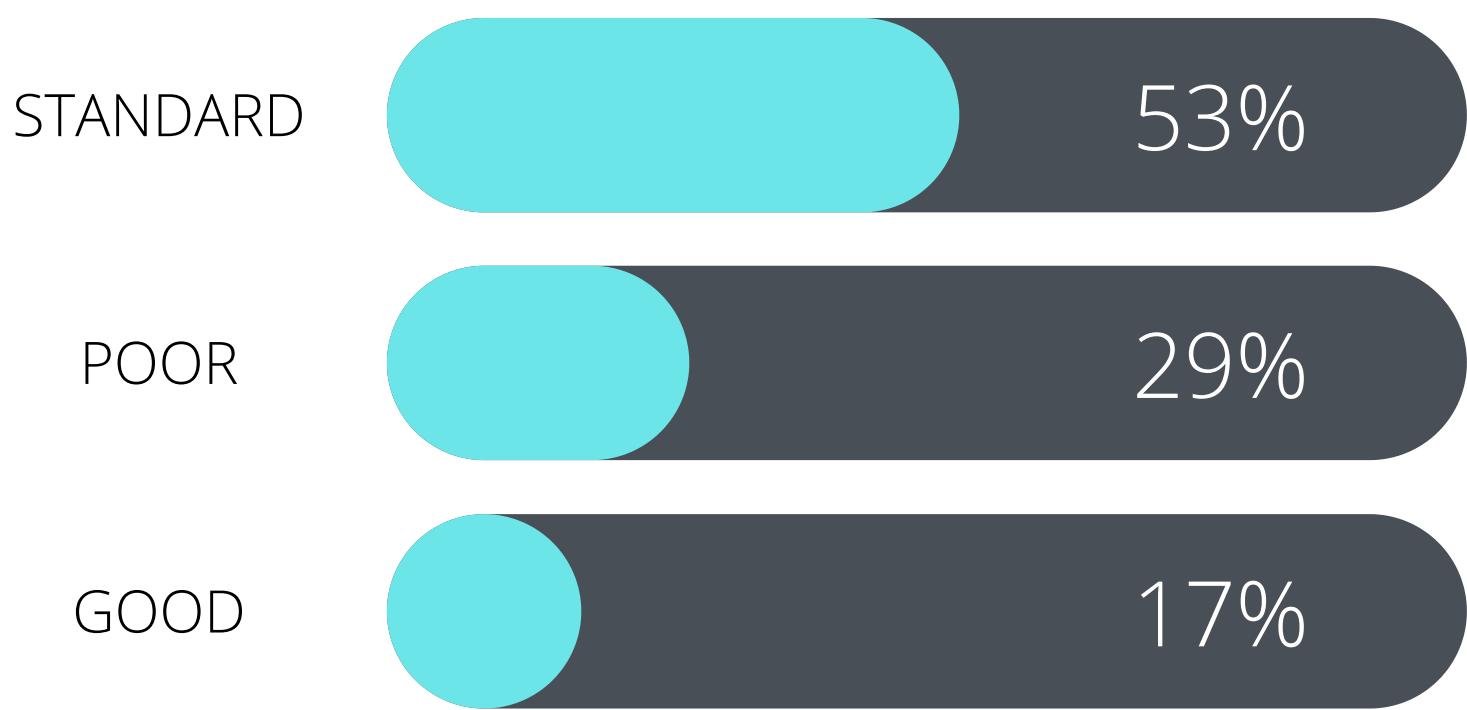
1

Resultados com a base de teste

BASE TREINO
100 MIL CLIENTES

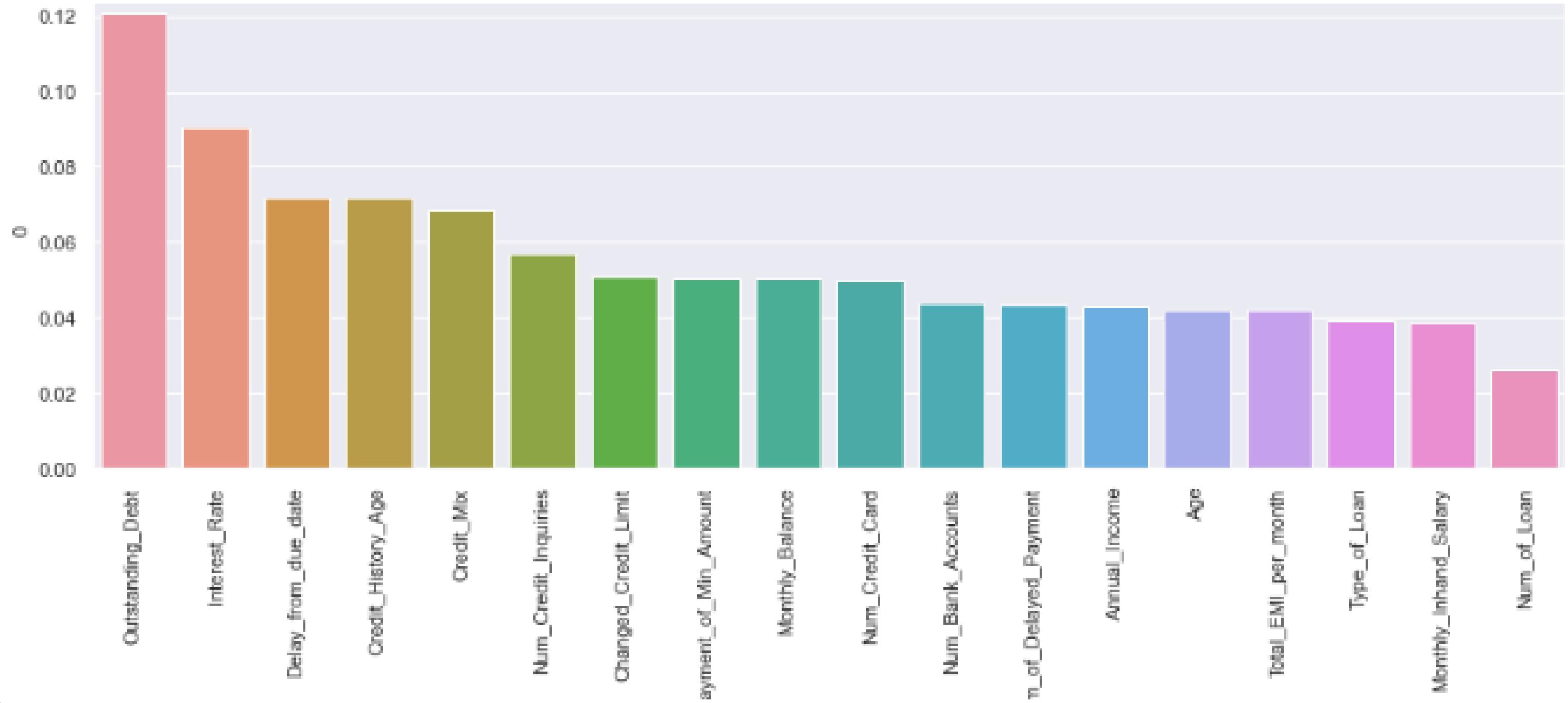


BASE TESTE
50 MIL CLIENTES



Importância das Features

Afinal, quais features mais influenciam o nosso modelo?



Próximos passos

- Entrega do modelo produtizado para a área de negócio;
- Acompanhamento dos resultados;
- Melhorias no modelo - Apresentação mensal.





credbank