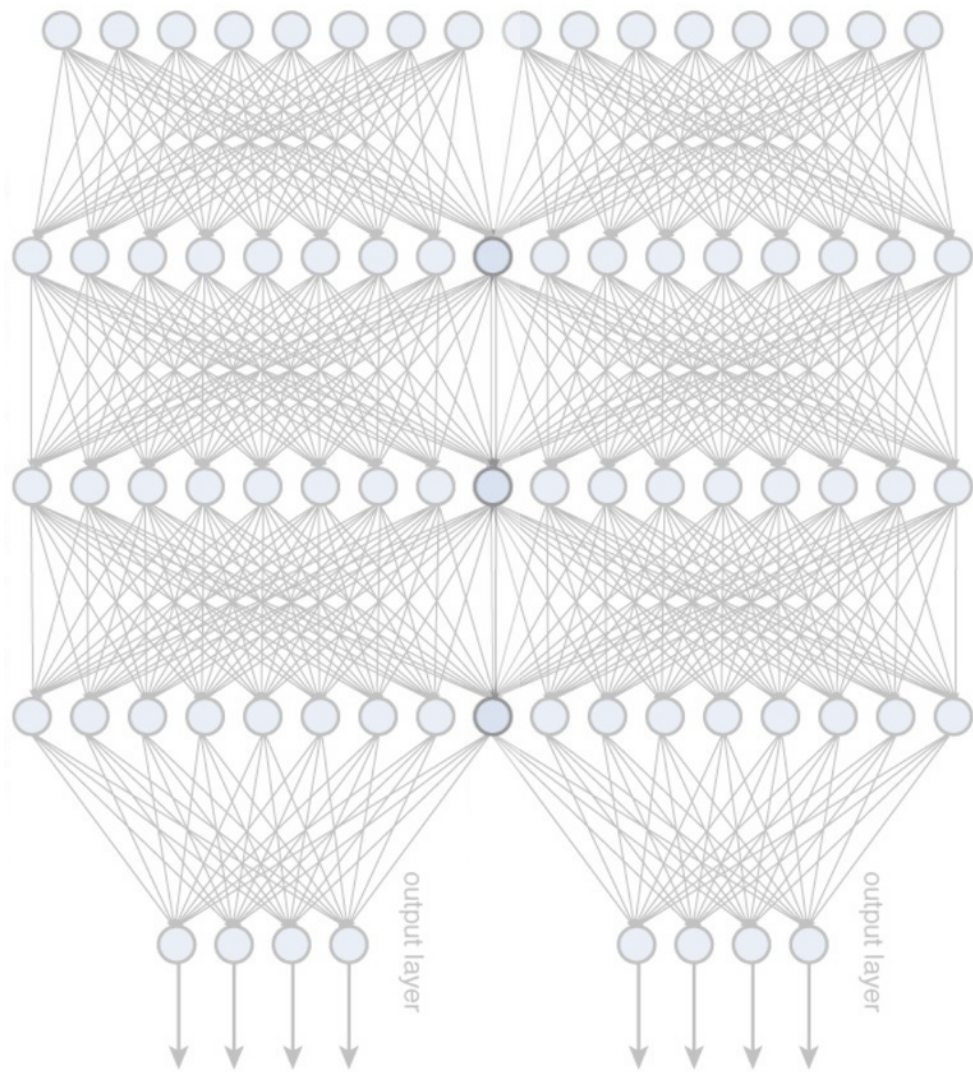




NEUROTECH

Processo seletivo

Modelo de Concessão de Crédito



Índice

- Projeto
- Perfil dos dados
- Análise exploratória
 - Análise distribuição geográfica
- Treinamento e otimização
- Análise financeira

Cientista de dados



Antônio Ramos

- Engenheiro de Produção
- Mestrando em Computação Inteligente
- Cientista de Dados
- Atuação em projeto envolvendo IA na área da saúde, indústria, moda, tecnologia, e outros.



Projeto

Projeto

Objetivo: Desenvolver um modelo de concessão de crédito.

Etapas:

1. Entendimento da base e análise exploratória dos dados.
2. Pré-processamento das variáveis.
3. Treinamento de um modelo de classificação binária.
4. Análise técnica da performance do modelo, medida sobre a base de Teste.
5. Análise financeira do modelo. Para este ponto, observe a subseção `Análise financeira`.
6. Escoragem da base Out-of-time, para posterior avaliação da performance nessa base com o alvo detido pelo Prophet/Neurolake.

Entregáveis:

1. Um Jupyter Notebook legível, contendo as etapas do projeto;
2. Apresentação do projeto;
3. Base Out-of-time com previsões.



Perfil dos dados

Perfil dos dados

Dados de treinamento

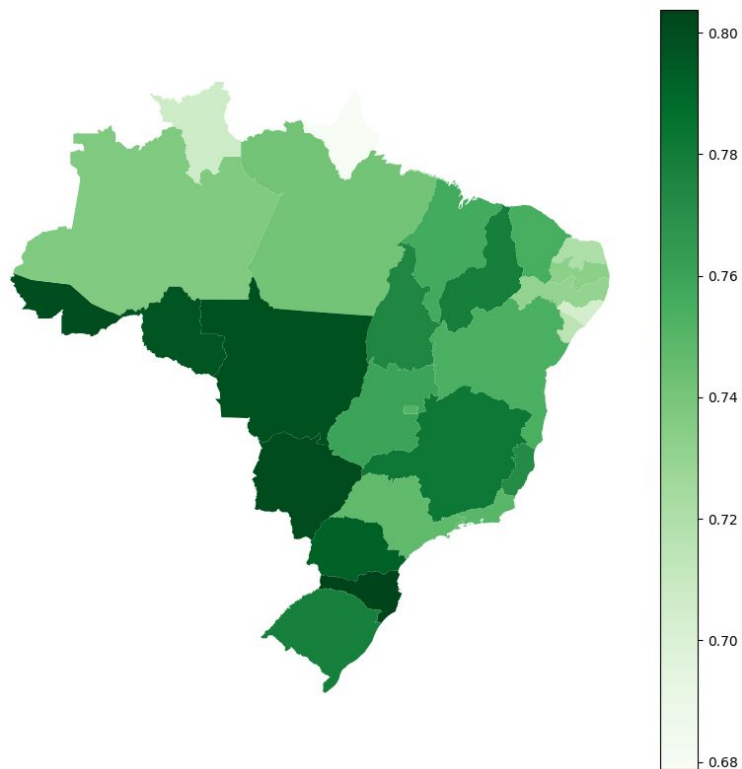
- **Tamanho da base (linhas,colunas):** (120.750, 151)
- **Início da base:** 02/01/2017
- **Fim da base:** 31/08/2017
- **Clientes únicos:** 120.750
- **0 (Bom pagador):** 91.163 / 75%
- **1 (Mau pagador):** 29.587 / 25%



Análise exploratória

Análise distribuição geográfica

Análise distribuição geográfica



A análise teve como foco entender como a propensão de ser um bom pagador varia entre diferentes regiões geográficas.

É possível perceber que, geograficamente, os bons pagadores estão bem disseminados, o que é um ótimo sinal pois não corremos o risco de ter uma viés por parte da localização geográfica.

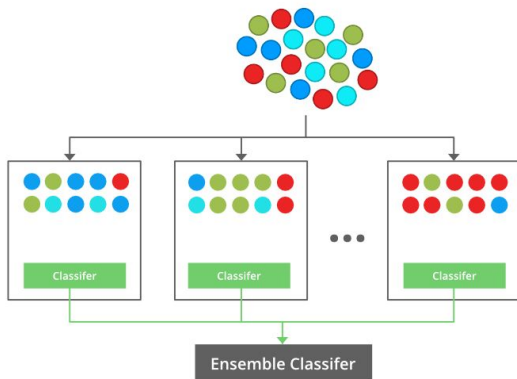
Porém vale ressaltar que temos alguns estados com maior concentração.

Treinamento e otimização

Treinamento e otimização

XGBClassifier

Escolhi o XGBoost pois ele atua com o bootstrapping, que é uma técnica de reamostragem com substituição. Isso permite que o modelo não seja apenas robusto contra overfitting, mas também maneje a variância de uma maneira que melhore a generalização.



Otimização de hiperparâmetros

A otimização utilizou uma pesquisa de grade. Defini um conjunto de parâmetros a serem testados, incluindo o número de estimadores, a taxa de aprendizado e a profundidade máxima das árvores.

```
param_grid = {  
    'n_estimators': [100, 200, 300, 500, 1000],  
    'learning_rate': [0.001, 0.01, 0.05, 0.1, 0.3],  
    'max_depth': [3, 4, 5, 8, 10],  
}
```

Melhores parâmetros:

- learning_rate: 0.05
- max_depth: 5
- n_estimators: 500

Melhor AUC-ROC no conjunto de teste:

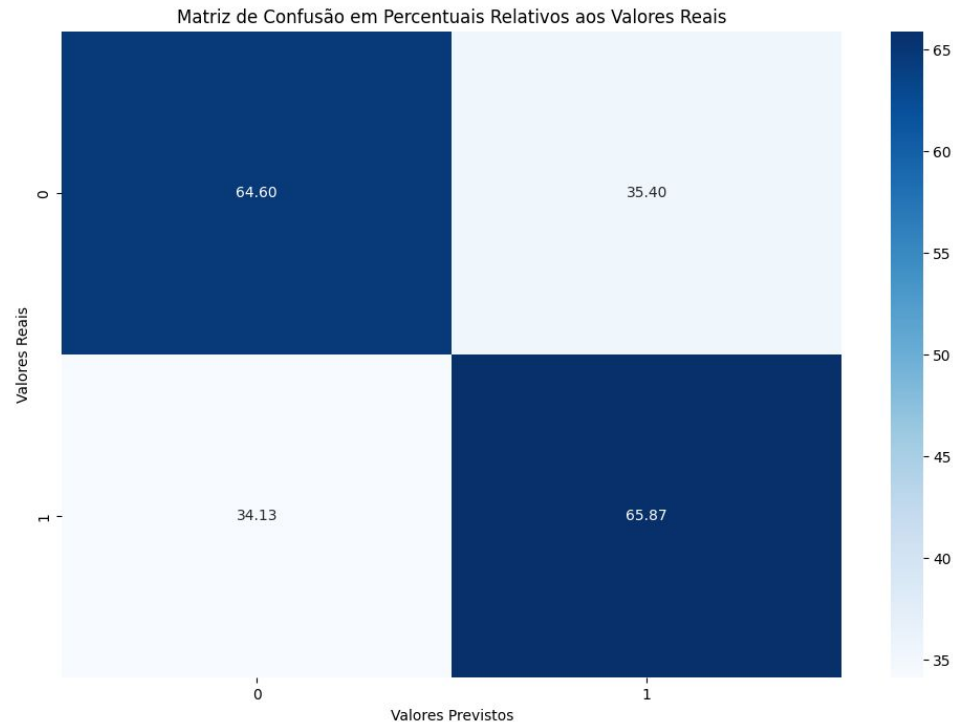
0.7067



Capacidade do modelo

O modelo acerta **64.60%** dos bom pagador e **65.87%** dos maus pagador. Isto indicar que não há um viés forte em favor de uma classe sobre a outra.

No entanto, a taxa de erro também é notável que será melhorada.



Análise financeira

Análise financeira

Política AS-IS

O ponto de corte são **clientes com 28 anos ou menos.**

Carteira de crédito aprovado para o mês de agosto de 2017 foi de **R\$5.432.000.**

Política TO-BE

O ponto de corte é o cliente ter uma **probabilidade igual ou maior a 59,28% (Mediana da probabilidade) de ser bom pagador.**

Carteira de crédito aprovado para o mês de agosto de 2017 foi de **R\$3.653.000.**

A política TO-BE pode estar contribuindo para uma carteira de maior qualidade ao evitar empréstimos a clientes com maior probabilidade de inadimplência. Porém, ainda é preciso melhorar a qualidade do modelo para tornar a regra mais robusta.





NEUROTECH