

Propósito

Este desafio tem como objetivo avaliar sua forma de resolver problemas, portanto, as análises e conclusões, estrutura e a qualidade do código, bem como as tecnologias e técnicas utilizadas vão dizer muito sobre como você é capaz de fazer isso!

Este desafio avaliará as seguintes habilidades:

- Entendimento do problema e capacidade de abstração
- Modelagem de dados utilizando algoritmos para métodos supervisionados
- Execução e avaliação do modelo criado
- Análise de estabilidade das variáveis
- Reprodutibilidade do modelo criado
- Estrutura e qualidade do código
- Forma da apresentação dos resultados

Sobre o Desafio

O desafio é a construção de um modelo de Credit Scoring cuja base de dados está localizada no [link](#). O arquivo contém 81 variáveis e 10.738 registros, sendo que o "id" é a chave única da tabela, "safra" é o mês ano de concessão do crédito, "y" é a variável *target* e as demais são variáveis preditoras mascaradas.

Seu trabalho consiste em:

1. Desenvolver um modelo para prever a variável *target*
2. Justificar o algoritmo e parâmetros utilizados
3. Apresentar métricas de performance do modelo
4. Explicar e apresentar os resultados

Stack de Tecnologias

- Linguagem: SAS, R, ou Python.

Não há restrições para uso de bibliotecas que possam otimizar o seu código mas esperamos que você use código que domine e seja capaz de entender e explicar.

Não basta apenas você saber executar o projeto

É importante que seu projeto tenha uma documentação consistente que permita a qualquer pessoa executar ou modificar o modelo criado.

O que esperamos ver ao final?

Nosso time está bastante curioso para ver e analisar seu problema. E queremos compartilhar com você alguns dos pontos que avaliaremos no seu projeto:

1. **Interpretação e metodologia:** mostrar de forma clara, por meio da estrutura do código e da documentação, a estratégia adotada para a resolução do problema, bem como todas as premissas assumidas e suas razões.
2. **Modelagem:** explicação da escolha do algoritmo, pontos fracos e fortes perante aos dados e também as técnicas.
3. **Performance:** Apresentar métricas de avaliação do modelo. O modelo é bom e separa bem?
4. **Manutenibilidade:** é um código legível e de fácil manutenção. Segue premissas claras de padronização de código.
5. **Visualização:** Apresentação dos resultados de forma clara e de fácil entendimento através de gráficos, tabelas, web applications, relatórios, entre outros.

O tempo estimado de criação desse projeto é de 48 horas. **BOA SORTE!**