

Case - Modelo preditivo para aprovação de crédito

Modelos de score de crédito calculam a probabilidade de inadimplência e são uma das principais ferramentas utilizadas por diversas empresas para aprovar ou negar um crédito.

O objetivo deste desafio é criar um modelo preditivo calculando a probabilidade de inadimplência de cada novo pedido de crédito.

Dados

Cada linha representa um cliente e as colunas representam os dados (informações) desses clientes.

A variável resposta é a coluna inadimplente, que indica se o tomador veio a se tornar inadimplente(1) ou não(0).

As variáveis da base de dados são descritas abaixo:

- idade: A idade do cliente.
- numero_de_dependentes: O número de pessoas dependentes do cliente.
- salario_mensal: Salário mensal do cliente.
- numero_emprestimos_imobiliarios: Quantidade de empréstimos imobiliários que o cliente possui em aberto.
- numero_vezes_passou_90_dias: Número de vezes que o tomador passou mais de 90 dias em atraso.
- util_linhas_inseguras: Quanto que o cliente está usando, relativamente ao limite dele, de linhas de crédito que não são seguradas por qualquer bem do tomador e.g: imóveis, carros etc.
- vezes_passou_de_30_59_dias: Número de vezes que o cliente atrasou, entre 30 e 59 dias, o pagamento de um empréstimo.
- razao_debito: Razão entre as dívidas e o patrimônio do tomador. $\text{razão débito} = \text{Dividas} / \text{Patrimônio}$
- numero_linhas_crdto_aberto: Número de empréstimos em aberto pelo cliente.
- numero_de_vezes_que_passou_60_89_dias: Número de vezes que o cliente atrasou, entre 60 e 89 dias, o pagamento de um empréstimo.
- score_1: score de risco de crédito bureau 1.
- score_2: score de risco de crédito bureau 2.
- score_3: score de risco de crédito bureau 3.
- score_4: score de risco de crédito bureau 4.

Obs: Estes dados foram retirados de terceiros, portanto é possível que existam incoerências, o que é perfeitamente comum em dados reais.

Objetivo

Construir um modelo preditivo utilizando o arquivo `case_treino.csv`.

Utilize este modelo para gerar as previsões na base `teste.csv`, inserindo uma nova coluna na tabela de dados do arquivo `teste.csv` que contenha as previsões e nomeie esta coluna com o nome "inadimplente".

Espera-se que não seja necessário mais que um dia para a resolução deste problema.

Avaliação

Serão avaliados o código, preferencialmente em python ou R, e o desempenho do modelo na base de teste (utilizaremos a métrica de AUC).

Para isto, só será necessário enviar o código com tudo o que foi feito e a base de dados de teste com a coluna de previsões.