



DESAFIO TÉCNICO - v1.2

CIÊNCIA DE DADOS

OBJETIVO

O objetivo deste desafio é permitir que o candidato demonstre seu conhecimento na preparação de dados, análise, criação e avaliação de modelos de previsão.

PROBLEMA

A *Acme* é uma empresa de varejo que vende diferentes itens para seus clientes. Para aumentar as suas vendas, ela decidiu conceder crédito para estes clientes com base em um modelo automático de risco. Mais especificamente, dadas diversas informações sobre um cliente *C*, obtidas no momento da requisição de crédito, ela quer inferir o risco associado a *C*. Entende-se como risco a probabilidade de que *C* não vai pagar a dívida que ele contraiu nos meses subsequentes, se tornando assim um mau pagador.

Para tanto, a *Acme* coletou dados de clientes solicitantes durante os meses de agosto e setembro de 2015, registrando o seu desempenho como bons ou maus pagadores nos meses subsequentes. Em particular, para cada cliente, a empresa coletou os seguintes dados:

- A data de solicitação de crédito (data de referência);
- Informação demográfica como idade, sexo etc;
- Informação sobre emprego do cliente;
- Informação sobre renda;
- Informação sobre o histórico de débitos passados.

Cada cliente foi representado por 15 colunas de dados, descritas abaixo:

Nome da Variável	Descrição	Tipo
refdate	Data de referência (forma YYYY-MM-DD)	Data
demo_age	Idade	Numérica
demo_gender	Gênero	Categórica
demo_marital_status	Estado civil	Categórica

demo_schooling	Escolaridade	Categórica
demo_residence_time	Tempo de residência	Numérica
demo_residence_kind	Tipo de residência	Categórica
job_formal	Tem emprego formal?	Lógica
job_retired	Aposentado?	Lógica
job_employment_time	Tempo empregado	Numérica
job_kind	Tipo de emprego	Categórica
job_employment_experience	Experiência em emprego	Categórica
income_social_benefits	Recebe auxílio social?	Lógica
income_salary	Salário declarado	Numérica
debt_profile	Perfil como pagador	Categórica
debt_prev_history	Já foi mau pagador no mercado?	Lógica
debt_bureau_score	Escore de crédito no mercado (NaN = desconhecido)	Numérica
target	Mau pagador? (1 se mau pagador)	Lógica

O desempenho do cliente, como bom ou mau pagador, é registrado na coluna **target**. Em particular, um cliente com **target** = 1 é considerado mau pagador.

Os dados descritos nesta seção serão fornecidos em anexo com este enunciado.

REQUISITOS

1. A solução proposta deve ser construída em um notebook Jupyter, usando a pilha de aprendizagem de máquina do Python (sklearn e bibliotecas associadas) para construção do modelo proposto.
2. Os dados de agosto de 2015 (cerca de 10000 clientes) devem ser usados para treinar o modelo enquanto os de setembro de 2015 (cerca de 5000 clientes), devem ser usados para avaliar o desempenho do modelo.
3. O notebook deve incluir todo o processo de análise, preparação e transformação de dados que for realizado.
4. Embora possam ser usados quaisquer algoritmos e técnicas de preparação de dados e previsão que o candidato considerar adequadas, espera-se que minimamente sejam testados *ensembles* (ex: florestas aleatórias ou métodos de boosting).

5. Note que para a empresa Acme é fundamental o reconhecimento de maus pagadores, que são a *minoría* da população. Portanto, uma alta acurácia não é necessariamente a forma mais adequada de medir o desempenho do método proposto.
6. Ainda como parte da avaliação, deve ser reportada a importância de cada um dos atributos usados no modelo proposto.

O DESAFIO

A sua missão é propor um modelo de previsão usando a pilha de aprendizagem de máquina do Python baseada no sklearn.

A ENTREGA

Após a criação do modelo, todos os códigos implementados e a avaliação do mesmo devem ser disponibilizados em um repositório público do github, através de um notebook Jupyter. No repositório, você pode incluir um arquivo README.MD para fornecer informações adicionais necessárias para a execução do seu notebook.

A AVALIAÇÃO

Nosso time técnico irá clonar o repositório e avaliar o modelo proposto. Depois disso, faremos uma reunião com você, onde você vai discorrer sobre as suas decisões técnicas e tirar as dúvidas que nossos avaliadores ainda possam ter. Note que serão bem vindas todas as informações no notebook que ajudem na compreensão de modelos e decisões tomadas.