



dev@meantrix.com

Desafio

Sugestões para o desafio:

- Responda ao desafio num prazo máximo de 5 dias;
- Envie sua resposta para dev@meantrix.com com o título: desafio-seu nome ;
- Sua resposta deverá conter um arquivo .py com os códigos e comentários (se necessários) dessa linguagem, um arquivo .Rmd com os códigos, funções gráficas e conclusões feitos a partir da linguagem R e demais subrotinas criadas todas compactadas em um único arquivo .zip;
- Procure utilizar as melhores práticas de programação e justificar sua análise se achar necessário;

Contextualização

O base de dados [fictícia] em formato CSV representa um conjunto de funcionários de uma determinada empresa. Há um grande número de informações, como sexo, idade, salário, estado civil, número de promoções etc.

A seguir temos a explicação de algumas variáveis:

Attrition: funcionário desligado (essa é a variável *target*)

Education

1 'Ensino Médio' 2 'Técnico' 3 'Graduação' 4 'Mestrado' 5 'Doutorado'

EnvironmentSatisfaction

1 'Baixo' 2 'Médio' 3 'Alto' 4 'Muito alto'

JobInvolvement

1 'Baixo' 2 'Médio' 3 'Alto' 4 'Muito alto'

JobSatisfaction

1 'Baixo' 2 'Médio' 3 'Alto' 4 'Muito alto'

PerformanceRating

1 'Baixa' 2 'Boa' 3 'Excelente' 4 'Extraordinária'

RelationshipSatisfaction

1 'Baixo' 2 'Médio' 3 'Alto' 4 'Muito alto'

WorkLifeBalance

1 'Ruim' 2 'Boa' 3 'Muito boa' 4 'Ótima'

A maioria das demais variáveis são auto-explicativas, mas algumas não possuem uma definição clara. Cabe a você decidir utilizá-las ou não!

O problema:

Desejamos criar um modelo capaz de prever quais funcionários pedirão para ser desligados (coluna 'attrition'), com base nas diversas variáveis disponíveis.

Em Python:

Aplique um algoritmo de ML (ou um conjunto deles) capaz de selecionar as principais variáveis e descobrir os funcionários que tem chance de serem desligados. Faça uma limpeza dos dados, retire variáveis redundantes, crie uma dataset de testes, etc, e tente mostrar um algoritmo funcional (se não conseguir ou se não for possível, explique porquê), e métricas de performance.

Em R:

Faça uma análise exploratória dos resultados acrescentando as demais variáveis contidas no dataset. Quais insights você pode obter a partir desses dados? Quais as características mais comuns e/ou mais importantes para o desligamento de um funcionário?

Desenvolva sua resposta em um arquivo .Rmd(Rmarkdown) criando um "mini-relatório".

Boa Sorte!