Teste Técnico da seleção de Ciência de Dados do Portal de Telemedicina

O arquivo `test\_data\_CANDIDATE.csv` contém alguns dados do paciente. O teste consiste em:

1) Analisando os dados

2) Construir um SEX PREDICTOR juntamente com um script que permita o uso do modelo.

\*\* Descrições de dados:

- idade em anos

- sexo: (M = masculino; F = feminino)

- cp: tipo de dor no peito

- trestbps: pressão arterial em repouso (em mm Hg na admissão ao hospital)

- col: colesterol sérico em mg/dl

- fbs: (açúcar no sangue em jejum > 120 mg/dl) (1 = verdadeiro; 0 = falso)

- restecg: resultados eletrocardiográficos em repouso

- talach: frequência cardíaca máxima alcançada

- nar: número de braços

- exang: angina induzida por exercício (1 = sim; 0 = não)

- oldpeak: depressão do segmento ST induzida pelo exercício em relação ao repouso

- inclinação: a inclinação do pico do segmento ST do exercício

- hc: cor do cabelo do paciente

- sk: cor da pele do paciente

- trf: tempo gasto no trânsito diariamente (em segundos)

- ca: número de vasos principais (0-3) coloridos por fluoroscopia

- tal: 3 = normal; 6 = defeito corrigido; 7 = defeito reversível

- age: in years

- sex: (M = male; F = female)

- cp: chest pain type

- trestbps: resting blood pressure (in mm Hg on admission to the hospital)

- chol: serum cholesterol in mg/dl

- fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

- restecg: resting electrocardiographic results

- thalach: maximum heart rate achieved

- nar: number of arms

- exang: exercise induced angina (1 = yes; 0 = no)

- oldpeak: ST depression induced by exercise relative to rest

- slope: the slope of the peak exercise ST segment

- hc: patient's hair colour

- sk: patient's skin colour

- trf: time spent in traffic daily (in seconds)

- ca: number of major vessels (0-3) colored by flourosopy

- thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

\*\* Entregáveis:

1) A análise dos dados

- Pode ser apresentado de qualquer forma, mas é preferível um notebook python;

- Uma análise e/ou breve explicação do modelo entregue na entrega 2 também deve estar presente.

2) Um modelo que prevê se um lote de pacientes é do sexo masculino ou feminino, dado um arquivo CSV totalmente novo contendo as mesmas colunas, exceto a coluna 'sexo'.

- O script deve ser um arquivo python chamado 'sex\_predictor.py'

- Espera-se um arquivo 'requirements.txt', para que as dependências necessárias possam ser instaladas em um ambiente virtual usando o comando 'pip install requirements.txt'. Não se esqueça de adicionar as dependências necessárias para a primeira entrega também.

- O script será usado desta forma:

(env) python sex\_predictor.py --input\_file newsample.csv

- O arquivo newsample.csv é um arquivo com a mesma estrutura do 'test\_data\_CANDIDATE.csv', mas com um número desconhecido de linhas (n)

- A saída esperada é uma tabela salva no mesmo diretório, mas com o nome 'newsample\_PREDICTIONS\_{candidate\_fullname}.csv' contendo apenas a coluna 'sex' que contém os valores previstos.

- Python versão 3.6.x é o preferido.

\*\*\* Observações

- É muito importante que descrevas cuidadosamente a tua linha de pensamento, os passos dados, as decisões que tomaste e porquê, para que possamos perceber a tua forma de trabalhar.

- Não se esqueça de adicionar um arquivo README detalhando como configurar e executar seu teste;

- Se tiver algum teste que falhou e decidiu abortar, mencione-o, os parâmetros utilizados e os resultados obtidos no seu relatório.

- Mostre-nos todos os modelos que testou e diga-nos porque escolheu aquele que nos apresentou.

\*\*\* Submissão

- O desafio deve ser entregue usando [GitHub](http://github.com/), [Bitbucket](http://bitbucket.org/) ou [GitLab](http://gitlab.com/). Compartilhe conosco e envie a URL do seu código por e-mail.

- Quaisquer perguntas devem ser feitas em resposta a este e-mail.

Boa sorte!