

Universidade de Brasília
Departamento de Ciência da Computação
Projeto 3, Introdução à Inteligência Artificial, Turma A, 1/2020
Prof. Díbio

O Hospital Israelita Albert Einstein, São Paulo (Brasil), coletou, organizou e disponibilizou dados de 5644 casos de COVID-19, com vários exames reais de saúde relacionados aos pacientes.

Em

<https://www.kaggle.com/dataset/e626783d4672f182e7870b1bbe75fae66bdfb232289da0a61f08c2ceb01cab01> estão disponibilizados esses dados de forma anônima, com dados clínicos normalizados (média zero, desvio padrão 1).

O projeto 3 consiste em desenvolver e aplicar um código de “Random Forest” nesses dados, e prever/responder as seguintes questões:

1. Baseando-se nos dados de laboratório (sem PCR) é possível prever se o paciente estava positivo ou negativo para COVID-19? Com qual precisão? Quais os 10 (dez) testes/variáveis que são mais relevantes nessa resposta?
2. Baseando-se nos dados de laboratório é possível prever quais casos, e por conta de quais variáveis/testes os pacientes teriam que ser: 1) internados em enfermaria, 2) internados em unidade semi-intensiva, 3) internados em unidade intensiva, 4) acompanhados em casa?

Um mini-relatório de até (máximo) 5 páginas deve ser enviado explicando sua solução e comentando os resultados, juntamente com o código completo e arquivos explicativos.

O código deve ser bem documentado, pode ser escrito nas linguagens (C, C++ ou Python), de forma modular com funções para cada tarefa independente, realizado por até dois (2) estudantes do curso usando “pair programming”, e entregue via sistema <http://aprender3.unb.br> do curso, no prazo estipulado.