

## Enunciado Trabalho em Grupos 2

O objetivo deste trabalho é explorar e treinar modelos de aprendizado sobre um conjunto de dados dado. Ao final, será entregue um notebook Python que faz toda a análise exploratória sobre o conjunto de dados dado, treina adequadamente um modelo de aprendizado supervisionado considerado o melhor e, ao final, é capaz de ler um arquivo para teste e avaliar seu desempenho.

O trabalho é feito pelos mesmos grupos do Trabalho em Grupos 1.

### Conjunto de dados: treino.csv

Inspirado em um naufrágio relativamente recente ocorrido no Mediterrâneo, o conjunto de dados apresenta 7.124 registros contendo passageiros do navio e com o rótulo de sobrevivente, ou não. Os atributos estão separados por vírgulas, valores textuais não estão entre aspas e tem alguns nomes que usam acentos. Atributos:

- IDpassageiro – identificação única por passageiro, na forma XXXX\_YY, onde XXXX é o número do grupo (tipicamente família) e YY é o número do integrante no grupo;
- Nome – nome do passageiro (fictício)
- Deck, NumeroCabine e PosicaoCabine – identificação da cabine ocupada por deck, número e posição (se é do lado que o navio ancora nos portos (P) ou o lado oposto (S));
- Procedência – de onde veio o passageiro ao embarcar: de algum país da Europa, da América, da África ou de outro lugar;
- Destino – porto de desembarque do passageiro;
- Idade – idade do passageiro;
- VIP – se o passageiro adquiriu o pacote VIP;
- AllInclusive – se o passageiro optou pela modalidade de hospedagem all inclusive (tudo incluído);
- ServicoCabine, Restaurante, Shopping, SPA e ParquedeAguas – quanto o passageiro já dispendeu nesses serviços;
- Sobrevivente – atributo classe: anotação sobre se o passageiro sobreviveu, ou não, ao naufrágio.

O notebook Python deve executar corretamente no colab.research.google.com, considerando que o arquivo treino.csv esteja colocado no menu Files (aba vertical).

### O que é para fazer?

Elaborem um notebook Jupyter que deve conter:

- Análise exploratória do conjunto de dados
- Treinamento dos modelos de aprendizado, de acordo com os algoritmos escolhidos, com os ajustes dos hiper-parâmetros
- Seleção do algoritmo que melhor se ajustou aos dados
- Documentação do experimento de aprendizado feito, com acurácia, recall e F-score, e plotagem da curva ROC.
- Células ao final do notebook que leiam um arquivo teste.csv (não fornecido) e que avaliem igualmente o desempenho do modelo escolhido (acurácia, recall, F-score, curva ROC)

Além da programação, o notebook deve ter:

- Matrícula e nome dos alunos do grupo
- Células de documentação explicando em detalhes cada passo do código escrito.
- O relato dos resultados de cada algoritmo executado (tabelas, gráficos, etc) e quais hiper-parâmetros foram ajustados e para quais valores.
- Análise e interpretação dos resultados, com escolha e treinamento do modelo escolhido como mais adequado.

O professor vai disponibilizar uma sala de entrega para os notebooks de vocês. Cada grupo entrega um único notebook.

Padrão para o nome do notebook a ser entregue: <integrante1>.ipynb ou <integrante1>\_<integrante2>.ipynb

Data de entrega: Ver na agenda no Moodle.

Todos os notebooks serão executados contendo treino.csv e teste.csv(não fornecido a vocês).