

ANTES DE COMEÇARMOS, AS REGRAS DO JOGO:

1. O seu trabalho deve ser enviado em um arquivo de extensão .py ou .ipynb. Você deve enviar um único arquivo para o todo o seu trabalho, por meio do LMS e também uma cópia para o email ttavaresm@gmail.com, com o título “TRABALHO ADA”. A cópia é solicitada para prevenirmos o caso de alguma falha no envio ao LMS.
 - a. Se você estiver fazendo o trabalho em grupo, cada membro do grupo deve enviar uma cópia do trabalho via LMS.
2. Na primeira linha (ou célula) do seu trabalho deve constar o seu nome completo.
 - a. Caso você esteja fazendo o trabalho em grupo, inclua os nomes de todas as pessoas do grupo de trabalho nas primeiras linhas (ou célula).
3. Leia com atenção todas as instruções. Por favor comente os principais passos das suas soluções, explicando sucintamente o que você está fazendo e por que você está fazendo.
4. Por favor tenha em mente que todos os códigos incluídos por você no trabalho serão analisados por arsenal técnico especializado com o objetivo de garantir que a sua entrega não será um grande amontoado de códigos oriundos de IAs generativas com pequenas modificações. Caso isto fique constatado, o trabalho pode ser desconsiderado e avaliado como não entregue.
5. O prazo de entrega é até as 23:59:59 do dia 02 de maio.

IMPORTANTE TER EM MENTE

1. Tenha em mente que quero avaliar o seu aprendizado quanto aos seguintes conceitos:
 - Utilização do Scikit-learn
 - KNNs
 - Random Forests
 - Métricas de Avaliação de Modelos de Classificação
 - Otimização de hiperparâmetros
2. Esteja totalmente à vontade para incluir no seu trabalho qualquer tipo de rotina que julgar interessante e que vá além dos pontos mencionados no item (1). Por exemplo, você pode incluir visualizações de gráficos, pré-processamento de dados, validação cruzada, entre outros. Mas entenda que você será avaliado unicamente pelos pontos do item (1). Isto é, se estes pontos estiverem bem cobertos você já receberá a melhor avaliação possível. E pontos adicionais poderão render uma menção pelo capricho no trabalho, mas realmente não serão esperados ou cobrados.

O JOGO

1. Utilize o [Kaggle Datasets](#) para baixar um dataset tabular que represente um problema de classificação. Satisfeitas estas condições, você pode escolher o dataset que achar melhor.
2. Seu objetivo é treinar um KNN e uma RandomForest para resolver o problema de classificação proposto. Mas você deve necessariamente:
 - a. Utilizar o pacote scikit-learn para as principais etapas do trabalho
 - b. Otimizar os hiperparâmetros de cada um dos modelos usando necessariamente uma heurística como *grid search* ou *random search* (a aplicação do método escolhido para otimização de hiperparâmetros deve estar explícita no trabalho)
 - c. Para cada modelo, aferir as métricas de avaliação de modelos de classificação que achar mais adequadas (lembrando-se de usar sempre um conjunto de testes para avaliar a questão do overfitting)
3. Finalizados os treinamentos dos modelos com os respectivos hiperparâmetros otimizados (passo 2b), compare os resultados obtidos (no passo 2c) e informe qual modelo você entende que performou melhor, com base nas métricas de avaliação que estudamos e explicando como você interpretou estas métricas para chegar à conclusão que chegou.