

Aula 09 - Assíncrona - Atividade Prática Não Valorizada Sobre Tarefa de Classificação e Extração de Informações de Desempenho a Partir da Matriz de Confusão

Esta atividade corresponde à uma atividade **NÃO valorizada, mas que vale presença na Aula 09** do curso de *Machine Learning*.

Instruções:

- 1 - Reunam-se em grupos de **no máximo 3 pessoas**;
- 2 - A entrega deve ser realizada via CANVAS através de um artigo no formato PDF gerado a partir do *notebook*. Para tal fim, os grupos poderão utilizar o *Jupyter Notebook* ou o *Google Drive* para implementação do que foi solicitado.
- 3 - O período de disponibilidade para recebimento da atividade será de **31/08/23 às 17h, até 11/09/23, às 17h**.
- 4 - No cabeçalho do documento devem constar o **nome completo e o RA de TODOS os membros do grupo**.

Práticas a serem desenvolvidas:

- 1 - Criação de um *dataset* sintético para avaliação da tarefa de classificação;
- 2 - Divisão de dados em treinamento e teste;
- 3 - Criação e aplicação de um modelo de *Machine Learning*;
- 4 - Verificação da eficiência do modelo na tarefa de classificação através da Matriz de Confusão.

Exercício

Utilize a biblioteca `sklearn` para criar um *dataset* genérico a partir da função `make_classification`. O seu problema deve conter um total de 10000 amostras distribuídas em um conjunto de 10 *features*, onde você irá definir o número de *features* relevantes e informativas arbitrariamente, assim como valor do parâmetro `random_state`. O número de classes no problema **deve ser 2**. De posse do *dataset* criado, crie o código em Python para realizar as seguintes etapas:

- a)** Divida os dados de treinamento e teste na proporção 70% e 30%
- b)** Aplique o algoritmo SVM (Máquinas de Vetores Suporte - *Support Vector Machines*) para realizar a classificação.

Obs: O material da Aula Síncrona 01 apresenta um exemplo da aplicação deste algoritmo.

- c)** Pesquise sobre a função `classification_report` contida na biblioteca `sklearn.metrics` e utilize-a para obter as métricas de avaliação de desempenho do modelo que você implementou.