



MÓDULO 3 – APRENDIZADO DE MÁQUINA SUPERVISIONADO I

EXERCÍCIO 3

ÁRVORES DE DECISÃO E FLORESTAS ALEATÓRIAS KYPHOSIS DATASET

1 Descrição do Problema

A **Cifose** [1] é definida como o aumento anormal da convexidade posterior da coluna vertebral. Enquanto a maioria dos casos de cifose são brandos e necessitam apenas de monitoramento, casos mais sérios podem se debilitantes, causando dor e desconforto, além de dificuldades respiratórias, irregularidades cardiovasculares e, em casos muito severos, diminuir significativamente o tempo de vida. Geralmente, os pacientes não respondem bem a tratamentos conservativos, sendo recomendado uma cirurgia para corrigir a curvatura da coluna.

Neste exercício iremos utilizar **Árvores de Decisão e Florestas Aleatórias para classificar o resultado de cirurgias de correção da coluna em crianças**. Utilizaremos o dataset Kyphosis que contém as seguintes anotações:

- **Kyphosis**: a presença ou ausência de cifose após a cirurgia;
- **Age**: a idade da criança, em meses;
- **Number**: número de vértebras envolvidas na operação;
- **Start**: o número da vértebra mais elevada que foi operada;

2 Tarefas

Neste exercício, pedimos que você:

1. Inspeccione os dados.
2. Treine uma árvore de decisão para classificar a presença de cifose após a cirurgia.
3. Treine também florestas aleatórias variando o número de árvores geradas.
4. Classifique os dados de teste.
5. Calcule a matrix de confusão, a acurácia balanceada e compare os modelos treinados.

3 Arquivos

Os arquivos disponíveis no Moodle são:

- *kyphosis_train.data*: dados de treinamento;
- *kyphosis_val.data*: dados de validação;

4 Referências

1. *Kyphosis*. Wikipedia. <https://en.wikipedia.org/wiki/Kyphosis>.