

## ESTUDO DE CASO

### Patologia cardíaca em crianças e jovens

O objetivo principal deste trabalho é aplicar o processo de KDD para **predição e extração de conhecimento** de patologias cardíacas em crianças e jovens.

O conjunto de dados é o que vem sendo utilizado nas aulas práticas do curso desde o início do semestre letivo. Assim, a descrição do problema e dos dados já se encontra no SIGAA.

**Um subconjunto dos dados originais foi separado para ser usado como avaliação externa no Kaggle.**

Lembrando que, por se tratar de um **curso de mineração de dados**, além dos dados serem reais, o objetivo do estudo de caso é **explorar todas as fases do processo de KDD** (e não somente gerar um monte de modelos para atingir acurácia altas). Desta forma, o estudo de caso por si só possui especificidades que demandam um **aprofundamento no entendimento do problema e um esforço adicional para a manipulação dos dados** antes da modelagem.

A avaliação do estudo de caso será realizada em duas etapas, descritas a seguir.

#### **Primeira etapa**

Submeter no SIGAA um **mini relatório** de até 2 (**duas**) páginas descritos de **forma concisa**:

- As **principais limpezas** nos dados que foram realizadas;
- As **principais transformações** nos dados que foram realizadas;
- **Quais algoritmos** foram utilizados para gerar os modelos;
- Como os modelos foram avaliados (*hold-out*, validação cruzada...);
- Se houve otimização de hiperparâmetros dos algoritmos, como isto foi feito;
- Os resultados obtidos utilizando as métricas: **Accuracy, Recall, Precision e F-measure**.
- Uma interpretação do **modelo que obteve o melhor resultado, segundo as suas métricas**;
- **Não** são necessários capa, resumo (*abstract*) e referências (**basta o seu nome e matrícula**).

**Observação importante:** as métricas de desempenho devem ser apresentadas em duas versões: **com e sem a variável SOPRO**.

#### **Segunda etapa**

Será disponibilizado no **Kaggle** um **novo conjunto** de testes (**sem a variável SOPRO**).

Vocês devem fazer a predição dos exemplos deste conjunto e subir no Kaggle, que irá fazer a avaliação.

**Apenas o arquivo das predições**, na mesma ordem dos exemplos, é que deve ser subido.

O link do Kaggle será disponibilizado oportunamente.

## **Avaliação final do estudo de caso**

A avaliação será **individual** e é **subjetiva** e serão avaliados os seguintes itens:

- Qualidade da escrita do relatório;
- Corretude da metodologia aplicada para gerar os resultados;
- Resultados obtidos;
- Comparação dos resultados reportados no relatório e os avaliados no Kaggle.

## **Datas das entregas**

Primeira parte – submissão no SIGAA: até as **12h do dia 26/11/2025** (final da aula da quarta-feira).

Segunda parte – submissão no Kaggle: até as **23:55h do dia 28/11/2025** (último dia de aula).

**NÃO SERÁ ACEITA, EM NENHUMA HIPÓTESE, ENTREGAS POR MEIO DO E-MAIL DO PROFESSOR!!!**