

Atividade de Mineração de Dados: 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 trabalho deverá ser realizado em grupo, com os grupos que já foram definidos anteriormente.

Os dados a serem utilizados neste trabalho foram coletados no Real Hospital Português (RHP), anonimizados com a aprovação do Comitê de Ética do RHP. O Comitê de Ética da Universidade do Porto, Portugal, também aprovou o uso desses dados para estudos acadêmicos.

As inscrições consistem em crianças e jovens entre 0 e 19 anos, com ou sem patologia cardíaca (variável NORMAL X ANORMAL). A tabela abaixo fornece uma breve descrição das variáveis.

ID	ID do paciente anônimo
Peso	peso do paciente
Altura	altura do paciente
IMC	índice de massa corporal
Atendimento	data da visita
DN	data de nascimento
IDADE	idade
Convenio	seguro de saúde
PULSOS	pulsção
PA SISTOLICA	pressão arterial sistólica (PAS)
PA DIASTOLICA	pressão arterial diastólica (PAD)
PPA	resultado PAS/PAD
NORMAL X ANORMAL	ausência ou presença de patologia
B2	tipo do segundo som cardíaco
SOPRO	tipo de sopro
FC	frequência cardíaca
HDA 1	histórico da doença 1
HDA2	histórico da doença 2 (outro histórico)
SEXO	sexo do paciente
MOTIVO1	primeiro motivo para ser encaminhado para a clínica de cardiologia
MOTIVO2	segundo motivo para ser encaminhado para a clínica de cardiologia

A variável PPA é calculada a partir da PA SISTOLICA e PA DIASTOLICA, de acordo com tabelas clínicas cardíacas que relacionam o sexo do paciente com a idade e as pressões arteriais (sistólica e diastólica). Para poder melhorar a qualidade da sua análise, é necessário que você busque intervalos normalmente usados em cardiologia pediátrica, usando por exemplo, para as variáveis IDADE e IMC.

O que entregar?

Um relatório técnico com as seguintes características (formatação baseada no modelo da SBC):

- Possuir **entre 8 e 10 páginas**;
- Tamanho de folha A4 com coluna única, 3,5 cm para margem superior, 2,5 cm para margem inferior e 3,0 cm para margens laterais, sem cabeçalhos ou rodapés. A fonte principal deve ser Times, tamanho nominal de 12 pontos, com 6 pontos de espaço antes de cada parágrafo;
- Os títulos das seções devem estar em negrito, 13pt, alinhados à esquerda. Deve haver um espaço extra de 12pt antes de cada título. A numeração das seções é opcional. O primeiro parágrafo de cada seção não deve ser recuado, enquanto as primeiras linhas dos parágrafos subsequentes devem ser recuadas em 1,27 cm. Os títulos das subseções devem estar em negrito, 12pt, alinhados à esquerda;
- **Não** são necessários capa, resumo (*abstract*) e referências;
- O relatório pode ser enriquecido com tabelas, figuras e gráficos. As legendas devem ser centralizadas. A legenda de tabela deve ser colocada antes da tabela, enquanto que a legenda das figuras e gráficos vêm abaixo.

Elementos mínimos que devem ser abordados no relatório técnico

- Problema tratado e objetivo do trabalho;
- Base de dados original;
- Qualidade dos dados;
- Limpeza dos dados;
- Transformação dos dados;
- Mineração dos dados;
- Resultados (tanto de predição quanto de conhecimento extraído);
- Metodologia empregada para avaliação dos resultados.

Instruções adicionais

- É obrigatório usar ao menos **Árvore de Decisão** para gerar os resultados. Outros métodos também podem ser explorados a fim de melhorar os resultados.
- Indicar o link para o Colab, Github ou algum outro lugar onde é possível acessar o código dos experimentos.

Submissão e apresentação

- O relatório deverá ser submetido pelo SIGAA **até as 12h do dia 03/12** (**não serão aceitas entregas por e-mail**);
- Uma apresentação oral do trabalho realizado deverá ser feita entre os dias **03/12 e 10/12**:
 - Cada apresentação será de no máximo **10 minutos com mais 10 minutos para perguntas**;
 - Todos os integrantes do grupo devem estar presentes no dia da apresentação;
 - Será sorteado alguém do grupo para fazer a apresentação do trabalho;
 - Posteriormente, será sorteado alguém do grupo para responder às possíveis questões sobre trabalho.

O que será avaliado?

- Atendimento à formatação do relatório e tempo de apresentação;
- Qualidade escrita do relatório técnico;
- Qualidade da apresentação oral e das respostas da arguição posterior;
- Resultados obtidos;
- Corretude da metodologia aplicada para gerar os resultados.