

Universidade Federal do Ceará - Campus Quixadá
QXD0010 – Estruturas de Dados
Prof. Fabio Dias

ATIVIDADE PAI - ÁRVORE DE DECISÃO

A implementação do programa descrito neste documento deve ser entregue até a meia-noite do dia **09/07/2023** pelo Moodle.

Leia atentamente as instruções abaixo.

Instruções:

- Esta atividade pode ser feito em **dupla** ou **individualmente** e deve ser implementado (preferencialmente) usando a linguagem de programação C++ (**Não aceitarei mais do que dois alunos por atividade**)
- Identifique o seu código-fonte colocando o **nome** e **matrícula** dos integrantes da dupla como comentário no início de seu código.
- Indente corretamente o seu código para facilitar o entendimento. **Trabalhos com códigos maus indentados sofrerão redução na nota.**
- Deve-se utilizar a sua implementação da árvore binária ou a que o professor disponibilizou. Altere essa ED para melhor atender as exigências da atividade.
- **Pode utilizar outras ED da linguagem C++, menos de árvore.**
- **Apenas pode utilizar bibliotecas padrão do C++. Está completamente proibido usar biblioteca particular.**
- Os programas-fonte devem estar devidamente organizados e documentados.
- Observação: Lembre-se de desalocar os endereços de memória alocados quando os mesmos não forem mais ser usados.
- Não teremos caso de teste!!!!!!
- **Observação: Qualquer indício de plágio resultará em nota ZERO para todos os envolvidos.**

1 Sistema de Auxílio no Processo com Árvore de Decisão

Um programa de auxílio no processo de decisão tem por finalidade auxiliar a tomada de decisão mediante a realização de algumas perguntas ao usuário do programa. O programa irá fazer varias perguntas ao usuário com cada pergunta contendo pelo menos duas respostas. Para cada resposta irá direcionar a uma nova pergunta ou a uma decisão.

Por exemplo, digamos que você deseja saber onde jantar e o programa possui a seguinte configuração (Figura 1):



Figura 1: Árvore de Decisão

Se o usuário responder não a primeira pergunta e sim na segunda a decisão sugerida ao usuário será “Santa Grelha”.

Veja outro exemplo abaixo (Figura 2):

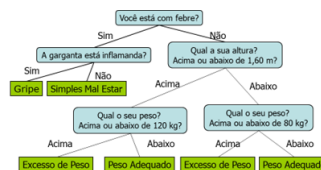


Figura 2: Árvore de Decisão.

O trabalho é desenvolver este programa de auxílio no processo de decisão.

1.1 Problema

Implemente em C++ um programa de auxílio a tomada de decisão para o diagnóstico de doença. O programa irá ler uma base de dados (dataset) onde teremos 41 doenças e 132 sintomas. Seu programa deve ler essa base de dados e gerar a árvore de decisão conforme Figura 2. Os nós internos serão os sintomas e as folhas serão o diagnóstico. Após isso, ele inicia o processo de diagnóstico, onde ele fará as perguntas com apenas respostas sendo Sim (1) ou Não (0). Ao final, ele deve informar o diagnóstico. Após isso, ele retorna ao início para começar outro diagnóstico.

Na Figura 3 temos o formato do dataset. As colunas representam os sintomas, sendo que a primeira coluna será a doença. Em cada linha teremos para cada sintomas 1 ou 0, indicando sim ou não para esse sintoma.

prognosis	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	muscle_wasting	vomiting	burning_micturition	spotting_urination	fatigue
Fungal infection	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Fungal infection	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
Allergy	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
GERD	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0
GERD	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0
GERD	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
GERD	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
GERD	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0
GERD	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0
GERD	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0

Figura 3: Data Set com os Sintomas e Doenças.

Observe que para uma mesma doença, podemos ter circunstância onde a resposta poderá ser sim (1) e/ou não (0) para um sintoma.

O dataset será disponibilizado em planilha como em .txt.

Possivelmente, ao final, poderemos ter um diagnóstico com duas ou mais doenças.

1.2 Arquivo TXT

O arquivo txt terá a seguinte formatação.

As primeiras 41 linhas serão strings contendo os nomes das doenças. Aqui, a posição da doença irá representar o código dessa doença. Por exemplo, a primeira doença é a AIDS, então o código da AIDS é 1, a segunda doença é a Acne, então o código dessa doença é o 2, e assim sucessivamente. Esse código será usado para identificar a doença.

As próximas 132 linhas serão strings contendo os nomes dos sintomas. De forma semelhante teremos o código para identificar os sintomas. O primeiro sintoma é o itching terá código 1, o segundo sintomas o skin rash código 2, e assim sucessivamente.

As próximas 4920 linhas serão linhas da relação doença sintomas, onde teremos apenas inteiros. Para cada linha na primeira coluna teremos o código da doença. A partir da segunda coluna, teremos 1 (Sim) ou 0 (Não) indicando se essa doença tem um sintoma. Na segunda coluna será o sintoma 1, na terceira coluna o sintoma 2, e assim sucessivamente.

Por exemplo, na Figura 4 temos três linhas exemplo da relação doença x sintomas. Aqui temos a doença de código 15, no caso Fungal infection. Na primeira relação, indica que teremos o sintoma 1, não tem o sintoma 2, tem o sintoma 3 e assim sucessivamente. Na segunda relação, indica que não teremos o sintoma 1, teremos o sintoma 2, teremos o sintoma 3 e assim sucessivamente.

15	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
15	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 4: Linhas exemplo da relação doença x sintomas.

1.3 O que deve ser submetido

- Deverá ser submetido:

- Os arquivos .h e .cpp da TAD Árvore Binária.
 - O(s) arquivo(s) contendo a implementação do programa.
 - O arquivo main.cpp contendo a função main() que será executado para o programa ser iniciado.
 - Os fonte devidamente organizados e documentados.
-
- Um dos parâmetros utilizados na avaliação da qualidade de uma implementação consiste na constatação da presença ou ausência de comentários. Comente o seu código. Mas também não comente por comentar, forneça bons comentários.
 - Não serão aceitos trabalhos submetidos após o prazo final.
 - Estou a disposição para quaisquer (+-) dúvidas, exceto dúvidas do tipo: "Não sei como fazer.". Essa irei responder com "Até próximo semestre!"