



Universidade Presbiteriana Mackenzie

ESTRUTURA DE DADOS II

Introdução à Ciência de Dados usando Árvores

Aplicação 2 (Apl2)

(Atividade em grupo: mínimo 3 e máximo 4 alunos)

Contextualização

O Brasil, assim como outros países, está comprometido com os Objetivos de Desenvolvimento Sustentável (ODS), uma agenda global que visa solucionar os principais desafios da humanidade, como erradicação da pobreza, saúde e bem-estar, educação de qualidade, entre outros. A fim de promover o uso consciente de dados públicos, este projeto visa utilizar *datasets* disponibilizados pela plataforma [Governo Aberto SP](#) para realizar uma análise exploratória de dados, aplicando os conceitos de árvores binárias de busca (BST) e árvores AVL.

Objetivos do Projeto

1. Escolher um *dataset* do portal [Governo Aberto SP](#) que esteja alinhado com algum dos ODS citados no [site oficial](#).
2. Modelar e estruturar os dados em uma árvore binária de busca (BST) e em uma árvore AVL.
3. Implementar operações sobre as árvores para realizar uma análise exploratória de dados.
4. Comparar o desempenho entre as duas implementações de árvore (BST e AVL) em termos de operações de busca, inserção e remoção.
5. Elaborar um relatório completo, detalhando o processo e os resultados obtidos.
6. Incluir gráficos para ilustrar dados e questões trabalhadas na análise exploratória e as conclusões obtidas.
7. Refletir sobre os desafios enfrentados e as lições aprendidas durante o desenvolvimento do projeto.

Etapas do Projeto

1. Escolha do *dataset*

- Visite o portal [Governo Aberto SP](#) e explore os *datasets* disponíveis.
- Escolha um *dataset* que esteja relacionado a uma ODS específica (por exemplo, "Saúde e Bem-Estar", "Educação de Qualidade" etc).

- Justifique a escolha do *dataset* e com qual ODS ele está alinhado.

2. Modelagem dos Dados

- Planeje como organizar os dados lidos do *dataset* (normalmente no formato CSV) em uma árvore binária. Cada nó da árvore deve conter as informações relevantes do *dataset*, como chaves e valores associados (como estatísticas, indicadores, etc.).
- Implemente dois tipos de árvores: uma **BST** e uma **árvore AVL**.
- Descreva como os dados serão inseridos e organizados na árvore.

3. Operações de Análise de Dados

- Desenvolva métodos que realizem as seguintes operações nas árvores:
 - **Inserção:** Inserir novos dados no formato de nós na árvore.
 - **Busca:** Realizar buscas por chaves específicas ou por intervalos de dados.
 - **Remoção:** Remover dados específicos solicitados pelo usuário.
 - **Análise Estatística:** Operações que permitam responder a perguntas sobre o *dataset* (ex.: "*Quantos eventos ocorreram entre 2015 e 2020?*" ou "*Qual o maior valor registrado em um determinado período?*" etc). Vocês precisam analisar o *dataset* e formular questões que serão respondidas explorando os dados. Os métodos implementados devem ser usados para ajudar a responder as perguntas formuladas.

4. Comparação de Desempenho

- Para cada operação (inserção, busca e remoção), registre a quantidade de comparações realizadas em cada árvore ou faça uso de outras métricas de desempenho que julgar relevante, como tempo de execução, por exemplo.
- Realize testes para comparar o desempenho das operações entre a **BST** e a **AVL**.
- Use gráficos para mostrar as diferenças de desempenho.

5. Documentação e Análise

- Escreva um relatório detalhado, contendo:
 1. Uma introdução sobre o tema e a relevância dos ODS.
 2. Descrição do *dataset* escolhido e sua relação com o ODS selecionado com uma justificativa para a escolha.
 3. Como os dados foram modelados nas árvores.
 4. Descrição da implementação das estruturas de dados e as operações realizadas.
 5. Quais perguntas exploratórias foram respondidas e o que foi descoberto a partir da análise.
 6. Comparação de desempenho entre a BST e a AVL, com gráficos e discussão sobre as diferenças observadas. Discutir sobre a eficácia de cada implementação de árvore para o tipo de análise realizada.
 7. Conclusão do trabalho com uma discussão sobre as implicações dos resultados e as limitações do estudo.
 8. Referências utilizadas.

Exemplos de Perguntas Exploratórias

1. ODS 3: Saúde e Bem-Estar

- **Pergunta exploratória:** *"Com base nos dados de saúde pública, quais regiões do estado apresentam a maior incidência de doenças evitáveis nos últimos cinco anos? Como esses dados se distribuem por faixa etária?"*
- **Análise:** Explore padrões de distribuição de dados relacionados à saúde pública em diferentes regiões e compare com grupos etários.

2. ODS 4: Educação de Qualidade

- **Pergunta exploratória:** *"Há uma correlação entre o índice de evasão escolar e a qualidade de infraestrutura das escolas nas diferentes regiões?"*
- **Análise:** Verifique como os índices de infraestrutura escolar estão relacionados com a taxa de evasão escolar e analise o impacto disso.

3. ODS 5: Igualdade de Gênero

- **Pergunta exploratória:** *"Quais setores da economia têm o maior desequilíbrio de gênero em cargos de liderança? Existe uma tendência de aumento da participação feminina ao longo do tempo?"*
- **Análise:** Organize os dados sobre a distribuição de gêneros em diferentes setores econômicos e cargos de liderança, e identifique tendências.

4. ODS 6: Água Potável e Saneamento

- **Pergunta exploratória:** *"Em quais regiões o acesso ao saneamento básico é mais limitado e como isso se relaciona com a qualidade de vida da população local?"*
- **Análise:** Use dados sobre saneamento básico para identificar as regiões mais vulneráveis e correlacione com indicadores de saúde e qualidade de vida.

5. ODS 13: Ação contra a Mudança Global do Clima

- **Pergunta exploratória:** *"Quais regiões do estado estão mais suscetíveis a desastres naturais causados por mudanças climáticas? Como a emissão de gases de efeito estufa tem variado nessas áreas?"*
- **Análise:** Use dados sobre desastres naturais e emissões de gases para identificar regiões vulneráveis e analisar as tendências de risco.

Entregas e Apresentação

O que entregar?

- Arquivos com a implementação completa da aplicação (versão BST e versão AVL).
 - Compacte o código-fonte (somente arquivos ***.java**) no **formato zip**.
 - No arquivo **zip**, inclua um **arquivo texto (.txt)** contendo a identificação do grupo.
- *Dataset* utilizado na aplicação.
 - Formato **csv**.
- Relatório, no formato **.docx** (*Word*), com os itens que foram pedidos na descrição do projeto.
- Apenas um aluno do grupo precisa realizar a entrega no Moodle.

O código-fonte será compilado com o compilador `javac` (21.0.2) na plataforma Windows da seguinte forma:

```
> javac *.java -encoding utf8
```

O código compilado será executado com `java` (21.0.2) na plataforma Windows da seguinte forma:

```
> java <Classe>
```

Sendo que `<Classe>` deve ser substituído pelo nome da classe que contém o método `public static void main(String[] args)`.

Apresentação

- Cada grupo deve preparar e apresentar o trabalho na data definida pelo professor (**obrigatório para nota – a não apresentação do trabalho implica em nota zero na avaliação para o integrante do grupo ausente**). A apresentação será feita para a sala dentro do tempo de 10 a 15 minutos (máximo). Todos os alunos do grupo precisam participar da apresentação e deixar claro sua contribuição no projeto.

Critérios de Avaliação do Projeto

A nota da atividade é calculada de acordo com os critérios da tabela a seguir.

Item de Avaliação	Pontuação Máxima	Parcial	Ausente
Escolha do <i>dataset</i> alinhado aos ODS	10 pontos	5 pontos	0 pontos
O <i>dataset</i> escolhido está alinhado e justificado adequadamente	Atende aos critérios	Relevância superficial	Ausente ou sem alinhamento
Modelagem e Organização dos Dados em Árvores (BST e AVL)	20 pontos	10 pontos	0 pontos

Item de Avaliação	Pontuação Máxima	Parcial	Ausente
Organização correta dos dados nas árvores, considerando a estrutura	Atende aos critérios	Parcialmente organizado	Organização inadequada
Implementação de Operações sobre a Árvore	20 pontos	10 pontos	0 pontos
Operações necessárias bem definidas e otimizadas	Atende aos critérios	Parcial ou com falhas	Não implementado
Comparação de Desempenho entre BST e AVL	15 pontos	7 pontos	0 pontos
Comparação clara e detalhada de desempenho	Atende aos critérios	Superficial ou com falhas	Ausente
Testes e Validação	10 pontos	5 pontos	0 pontos
Código testado e validado com exemplos claros	Atende aos critérios	Testes insuficientes	Ausente
Relatório Final	20 pontos	10 pontos	0 pontos
Relatório estruturado e com explicações detalhadas	Atende aos critérios	Faltam detalhes ou clareza	Ausente ou incompleto
Reflexão Final	5 pontos	2 pontos	0 pontos
Reflexão bem elaborada sobre os aprendizados e desafios	Atende aos critérios	Reflexão curta	Ausente

A tabela a seguir contém critérios de avaliação que podem reduzir a nota final da atividade.

Item Indesejável	Penalidade
Projeto é cópia de outro feito por outro grupo	Nota 0 (projetos são zerados)
O programa não compila ou possui erros graves	Nota 0 (projeto é zerado)
Ausência de referências no relatório	-1,0 ponto
Arquivos enviados em formato incorreto	-1,0 ponto
Relatório sem as seções obrigatórias (seguir o template)	-1,0 ponto por seção ausente
Código não documentado ou mal organizado	-1,0 ponto
Implementação incorreta da BST ou AVL	-2,0 pontos
O projeto não segue a proposta solicitada (ODS e análise)	-3,0 pontos
Testes não incluídos ou mal descritos	-2,0 pontos
Falta de gráficos ou análises no relatório final	-2,0 pontos
Relatório com menos de 3 páginas (muito superficial)	-1,5 pontos
Atraso na entrega (a menos que seja previamente acordado)	-1,5 ponto por dia

Template para o Relatório do Projeto Prático

1. Capa

- Título do Projeto
- Nome completo dos integrantes do grupo
- Universidade, Disciplina e Professor

2. Sumário

- Listagem das seções e subseções com números de página.

3. Introdução

- Breve apresentação do projeto.
- Objetivos gerais do projeto.
- Importância do tema abordado em relação aos ODS (Objetivos de Desenvolvimento Sustentável).

4. Escolha do *dataset*

- Descrição do *dataset* escolhido.
- Justificativa do alinhamento do *dataset* com um ou mais ODS.
- Link e fontes do *dataset* utilizado.

5. Modelagem dos Dados em Árvores

- Explicação sobre como os dados foram organizados e estruturados nas árvores (BST e AVL).
- Discussão sobre a escolha das chaves e a organização dos dados.
- Desenhos ou diagramas representando a estrutura da árvore.

6. Implementação das Operações sobre a Árvore

- Detalhes das operações implementadas para manipulação e consulta da árvore.
- Explicação de como as operações ajudam a responder às perguntas exploratórias.
- Trechos de código (opcional, se necessário para explicar a lógica).

7. Comparação de Desempenho entre BST e AVL

- Descrição do processo de comparação entre as duas estruturas.
- Métricas utilizadas (quantidade de comparações ou outra métrica).
- Tabelas e gráficos comparando os resultados.

8. Testes e Resultados

- Exemplos de entradas e saídas de dados.
- Descrição dos testes realizados para validar o código.

- Resultados obtidos nas operações e pesquisas.

9. Gráficos e Análises de Resultados

- Visualização dos dados (gráficos e tabelas).
- Análise dos resultados obtidos com base nas operações e perguntas exploratórias realizadas.

10. Conclusões

- Síntese das descobertas com base na análise exploratória.
- Avaliação do impacto da comparação de desempenho entre as estruturas (BST e AVL).
- Considerações finais sobre a análise realizada e seus impactos práticos.

11. Reflexão Final (individual, uma subseção por aluno(a))

- Breve reflexão pessoal sobre o aprendizado obtido durante a realização do projeto.
- Desafios enfrentados e como foram superados.

12. Referências

- Fontes e materiais consultados para a execução do projeto.
- Links para o *dataset*, artigos, livros, e outros materiais relevantes.