

Trabalho Prático: Busca e Ordenação
Disciplina: FGA0147 - Estruturas de Dados e Algoritmos
Prof. Dr. Nilton Correia da Silva
Faculdade UnB Gama - FGA
Universidade de Brasília - UnB

Aplicação de Algoritmos de Busca e Ordenação em Registros de Operações de Vendas da Amazon.com.

Fonte de dados: E-Commerce Sales Dataset



Gama, Distrito Federal

Sumário

1	Introdução	3
2	Dataset	4
2.1	Volumetria	4
2.2	Variáveis de Interesse	4
2.2.1	Categoria	4
2.2.2	Valor	4
3	Objetivo	5
3.1	Variáveis do Arquivo de Saída	5
3.2	Escolha do Método de Ordenação	5
4	Condições de Contorno da Solução	5
5	Entregáveis e Notas	5
5.1	Código-fonte da Solução - 5 pontos	6
5.2	Relatório Técnico - 5 pontos	6
5.2.1	Arquivo de saída - 2 pontos	6
5.2.2	Complexidade do Algoritmo de Ordenação escolhido - 3 pontos	6



Orientações

Trabalho Prático para composição da média final da disciplina de Estruturas de Dados e Algoritmos. Este trabalho foi elaborado considerando que sua resolução deve ser distribuída para um grupo de alunos.

A solução deverá ser em linguagem C ou C++. No caso do grupo adotar C++, poderá se valer de uma solução orientada a objetos, contudo sua solução não deve usar classes e contêineres já prontos da linguagem C++ - por exemplo, algoritmos de ordenação, busca, objetos vector, list, matrix, etc.



1 Introdução

Algoritmos de busca e ordenação são repetidamente demandados em diferentes tipos de análises de dados. Análises estatísticas que visam quantificar ou qualificar uma determinada característica, por vezes, requerem que o algoritmo realize ordenações e buscas em um dataset tendo como chave uma ou mais de suas variáveis - geralmente, as variáveis vêm em forma de colunas do dataset.

Neste trabalho teremos a oportunidade de exercitar algoritmos de ordenação e busca em um caso concreto representado por um dataset que foi preparado para que engenheiros de IA (Inteligência Artificial) e outros profissionais pudessem apresentar soluções para um problema envolvendo o referido dataset. Nosso objetivo aqui, contudo, não será apresentar uma solução final para o problema apresentado no link da fonte de dados (veja a referência abaixo do título do trabalho), mas sim, apresentar uma solução mais simples que será detalhada nas seções posteriores.

Caso o dataset disponibilizado tenha uma quantidade de registros maior do que os seus recursos computacionais conseguem processar, você pode excluir registros ou colunas (desnecessárias à solução), deixando apenas a quantidade máxima de dados que seus recursos computacionais conseguem processar.



2 Dataset

O dataset escolhido para este trabalho refere-se ao registro de vendas da empresa Amazon (arquivo **Amazon Sale Report.csv**). Ademais aos detalhes de compras, o dataset informa ainda dois campos de interesse aqui: Categoria e Valor.

Desafio para o Engenheiro de IA: Gerar um modelo de IA que seja capaz de inferir a lucratividade de comércios eletrônicos.

2.1 Volumetria

Instâncias: 128.974 linhas

Atributos: 24 colunas

Tamanho: Compactado: 6,5MB, Descompactado: 68,92 MB. Use apenas o arquivo **Amazon Sale Report.csv**

Formato: csv (valores separados por vírgula)

2.2 Variáveis de Interesse

Sua solução não demandará todas as variáveis do dataset. Veja abaixo o detalhamento das variáveis (colunas do dataset) que você precisará processar.

2.2.1 Categoria

Esta variável está na coluna **Category**. Esta coluna apresenta a categoria do produto vendido. Trata-se de um valor do tipo texto.

2.2.2 Valor

Esta variável está na coluna **Amount**. Esta coluna traz o valor da compra. Trata-se de um número com uma casa decimal.



3 Objetivo

A solução deste trabalho consiste em responder à seguinte questão:

Quais categorias de mercadorias promovem vendas com valores maiores?

Para responder a esta pergunta, você deve gerar um arquivo de saída contendo a categoria e o valor da compra.

Segue abaixo o detalhamento do arquivo de saída que sua solução deverá gerar.

3.1 Variáveis do Arquivo de Saída

Categoria:. Esta coluna deve trazer o tempo do voo em minutos. Você encontrará esta informação no campo *Category* do arquivo de entrada.

Valor:. Esta coluna deve trazer o valor da compra. Você encontrará esta informação no campo *Amount* do arquivo de entrada..

Este arquivo de estar em ordem decrescente pela coluna *Valor*.

3.2 Escolha do Método de Ordenação

A ordenação é um ponto importante deste trabalho. Um dos métodos abaixo deverá ser adotado para a solução:

1. Insert Sort
2. Bubble Sort
3. Selection Sort
4. Quick Sort

Veja: Principais Algoritmos de Ordenação

4 Condições de Contorno da Solução

Sua solução deve atender às seguintes condições:

1. O arquivo de entrada (dataset) deve ser lido por seu programa em sua apresentação original - não deve ser ordenado antes por outro programa.
2. As linhas do arquivo de saída devem estar em ordem decrescente pela coluna **Valor**.

5 Entregáveis e Notas

A avaliação deste trabalho se dará mediante a apresentação dos itens abaixo.



5.1 Código-fonte da Solução - 5 pontos

Mostrar execução do programa. Se for demorado executar para o dataset inteiro, prepare um dataset reduzido para esta etapa.

5.2 Relatório Técnico - 5 pontos

5.2.1 Arquivo de saída - 2 pontos

Mostrar o arquivo de saída gerado.

5.2.2 Complexidade do Algoritmo de Ordenação escolhido - 3 pontos

1. $O()$ - 1 ponto
2. Tempo de Processamento - 2 pontos

