

Educação Profissional Paulista

Técnico em
**Ciência de
Dados**

Lógica de programação e algoritmos

Decomposição de problemas

Aula 3

Código da aula: [DADOS]ANO1C3B3S19A3

Lógica de
programação e
algoritmos

Mapa da unidade 1 Componente 4

semana
21

Recursividade

semana
18

Prática de
decomposição de
problemas

semana
19

Você está aqui!

Decomposição de problemas

semana
23

Busca e
ordenação

Lógica de
programação e
algoritmos

Mapa da unidade 1 Componente 4

Você está aqui!

Decomposição de problemas

19

Aula 3

Código da aula: [DADOS]ANO1C3B3S19A3



Objetivo da aula

- Conhecer a importância de quebrar problemas complexos em partes menores e mais gerenciáveis, facilitando, assim, a análise e a solução.



Recursos didáticos

- Recurso audiovisual para exibição de vídeos e imagens;
- Acesso ao laboratório de informática e/ou internet.



Duração da aula

50 minutos.



Competências técnicas

- Usar técnicas para explorar e analisar dados, aplicar modelos estatísticos, identificar padrões, realizar inferências e tomar decisões baseadas em evidências;
- Compreender e dominar técnicas de manipulação de dados; extrair, transformar e carregar conjuntos de dados de diferentes fontes, garantindo a qualidade e a integridade dos dados; criar e compreender visualizações gráficas.



Competências socioemocionais

- Colaborar efetivamente com outros profissionais, como cientistas de dados e engenheiros de dados;
- trabalhar em equipes multifuncionais colaborando com colegas, gestores e clientes.

Construindo
o **conceito**

Passos básicos para aplicar a decomposição de problema em Ciência de Dados

Passos básicos para aplicar a decomposição de problema em Ciência de Dados

1. Identificação do problema

2. Decomposição em subproblemas

3. Coleta de dados

4. Análise exploratória

5. Pré-processamento de dados

6. Modelagem

7. Validação do modelo

8. Interpretação dos resultados

9. Comunicação dos resultados

Continuando a aplicação prática dos **passos básicos para aplicar a decomposição de problema em Ciência de Dados**, vamos abordar, agora, os tópicos de **Decomposição em subproblemas e Modelagem**, com o objetivo de mostrar o que esses passos diferem, dependendo da estratégia aplicada.

Elaborado especialmente para o curso.

Construindo o conceito

Aplicando os passos de decomposição de problema em Ciência de Dados para o estudo de caso do E-Commerce Boost

Quando aplicamos os **passos básicos para a decomposição de problema em Ciência de Dados** para o estudo de caso do *E-Commerce Boost*, percebemos que:

- Podemos trabalhar de formas diferentes para os tópicos de Decomposição em Subproblemas e Modelagem, dependendo da estratégia:

Tópico	Dividir e conquistar	Abordagem top-down	Abordagem bottom-up	Abstração
2 Decomposição em subproblemas	<p>Dividir o problema em etapas menores: coleta de dados, análise exploratória, pré-processamento, modelagem, validação e interpretação dos resultados.</p> <p>Identificar as variáveis relevantes: idade, gênero, histórico de compra e valor gasto.</p>	<p>Definir a pergunta principal: “Quais clientes têm maior probabilidade de realizar compras novamente?”.</p> <p>Identificar as variáveis relevantes: idade, gênero, histórico de compras e valor gasto.</p> <p>Dividir o problema em etapas menores: coleta de dados, análise exploratória, pré-processamento, modelagem, validação e interpretação dos resultados.</p>	<p>Analisar o histórico de compras de cada cliente individualmente.</p> <p>Identificar padrões de compra recorrentes e comportamentos de compra.</p> <p>Agrupar clientes com padrões de compra semelhantes para análise mais aprofundada.</p>	<p>Identificar as variáveis mais relevantes para a previsão de compras futuras.</p> <p>Simplificar o problema para focar apenas as variáveis-chave.</p> <p>Criar um modelo que utilize apenas as variáveis mais importantes para a previsão.</p>
6 Modelagem	<p>Para a estratégia de dividir e conquistar, a sugestão é utilizar o algoritmo de classificação <i>random forest</i>, devido à sua capacidade de lidar com conjuntos de dados complexos e de grande volume, bem como pela sua eficácia em lidar com problemas de classificação, como prever quais clientes têm maior probabilidade de realizar compras novamente.</p> <p>Além disso, o <i>random forest</i> é conhecido pela sua capacidade de lidar com variáveis de entrada de diferentes tipos e escalas, o que é útil quando se trabalha com dados de clientes que podem ter características diversas, como idade, gênero e histórico de compras, entre outras.</p> <p>Sua flexibilidade e capacidade de lidar com problemas complexos o tornam uma escolha adequada para a estratégia de dividir e conquistar, nesse contexto.</p>	<p>Para a estratégia <i>top-down</i>, a sugestão é utilizar o algoritmo de classificação regressão logística, por ser um modelo simples e interpretável, adequado para iniciar o processo de modelagem de forma mais direta e focada nos aspectos principais do problema.</p> <p>A regressão logística é fácil de interpretar e pode fornecer insights iniciais valiosos sobre o comportamento dos clientes. Isso a torna uma escolha adequada para iniciar o processo de modelagem com a estratégia <i>top-down</i>, focando os aspectos mais importantes do problema antes de considerar modelos mais complexos.</p>	<p>Para a estratégia <i>bottom-up</i>, a sugestão é utilizar um algoritmo de aprendizagem de máquina de classificação, como <i>random forest</i> ou <i>gradient boosting</i>, para prever as compras futuras dos clientes com base nos padrões identificados.</p> <p>Ambos os algoritmos têm a capacidade de lidar com dados complexos e de grande volume, o que os torna adequados para a estratégia de abordagem <i>bottom-up</i>, que começa com detalhes específicos e constrói uma solução mais ampla com base neles. A combinação desses algoritmos pode fornecer resultados precisos e robustos para o problema de prever quais clientes têm maior probabilidade de realizar compras novamente.</p>	<p>Para a estratégia de abstração, a sugestão é utilizar os algoritmos de classificação mais simples e interpretáveis, como a regressão logística ou mesmo árvores de decisão. Esses algoritmos são mais adequados quando se deseja simplificar um problema complexo, focando os aspectos mais importantes e ignorando os detalhes menos relevantes, conforme a definição de abstração.</p> <p>A regressão logística, é um modelo linear fácil de interpretar e pode fornecer insights diretos sobre quais variáveis estão mais relacionadas à probabilidade de um cliente realizar compras novamente. As árvores de decisão são modelos simples que dividem o conjunto de dados em subconjuntos menores com base em características específicas, o que pode ajudar a identificar padrões importantes nos dados, de forma mais abstrata.</p> <p>Portanto, para aplicar a estratégia de abstração, é mais apropriado escolher algoritmos mais simples e interpretáveis que possam ajudar a simplificar o problema e focar os aspectos mais relevantes para a previsão do comportamento dos clientes.</p>

Elaborado especialmente para o curso.

Colocando
em **prática**

Validação de modelo em Ciência de Dados



Durante a aula



Em grupo



ORGANIZEM-SE EM GRUPOS

Para entender como a **validação de modelo** é um dos passos básicos para a aplicação da decomposição de problema em Ciência de Dados, vocês irão realizar uma pesquisa na Internet.



PESQUISEM SOBRE O TEMA

Abram o navegador de internet e acessem o site de busca (Google, Bing, etc.). Na barra de pesquisa, digite "**Como fazer validação de modelo em Ciência de Dados**".



ANALISEM AS INFORMAÇÕES

- Quais são os métodos comuns de validação de modelo em Ciência de Dados?
- Como dividir os dados em conjuntos de treinamento e teste para validar um modelo?
- Quais são as métricas de avaliação de modelo mais utilizadas na validação?



ELABOREM UM RESUMO DA PESQUISA

Façam um resumo da pesquisa.

Ser
sempre +

Situação

Imagine que você foi contratado por uma empresa de comércio eletrônico. A empresa deseja aumentar suas vendas on-line e está buscando formas de entender melhor o comportamento de seus clientes para oferecer produtos mais personalizados.

Sua tarefa é analisar os dados de vendas e criar um modelo para prever quais produtos têm maior probabilidade de serem comprados por diferentes tipos de clientes.

Situação fictícia elaborada especialmente para o curso.

Ser
sempre +

Ação

Considerando a situação apresentada, organizem-se em grupos e discutam como resolver a tarefa encomendada pela empresa.

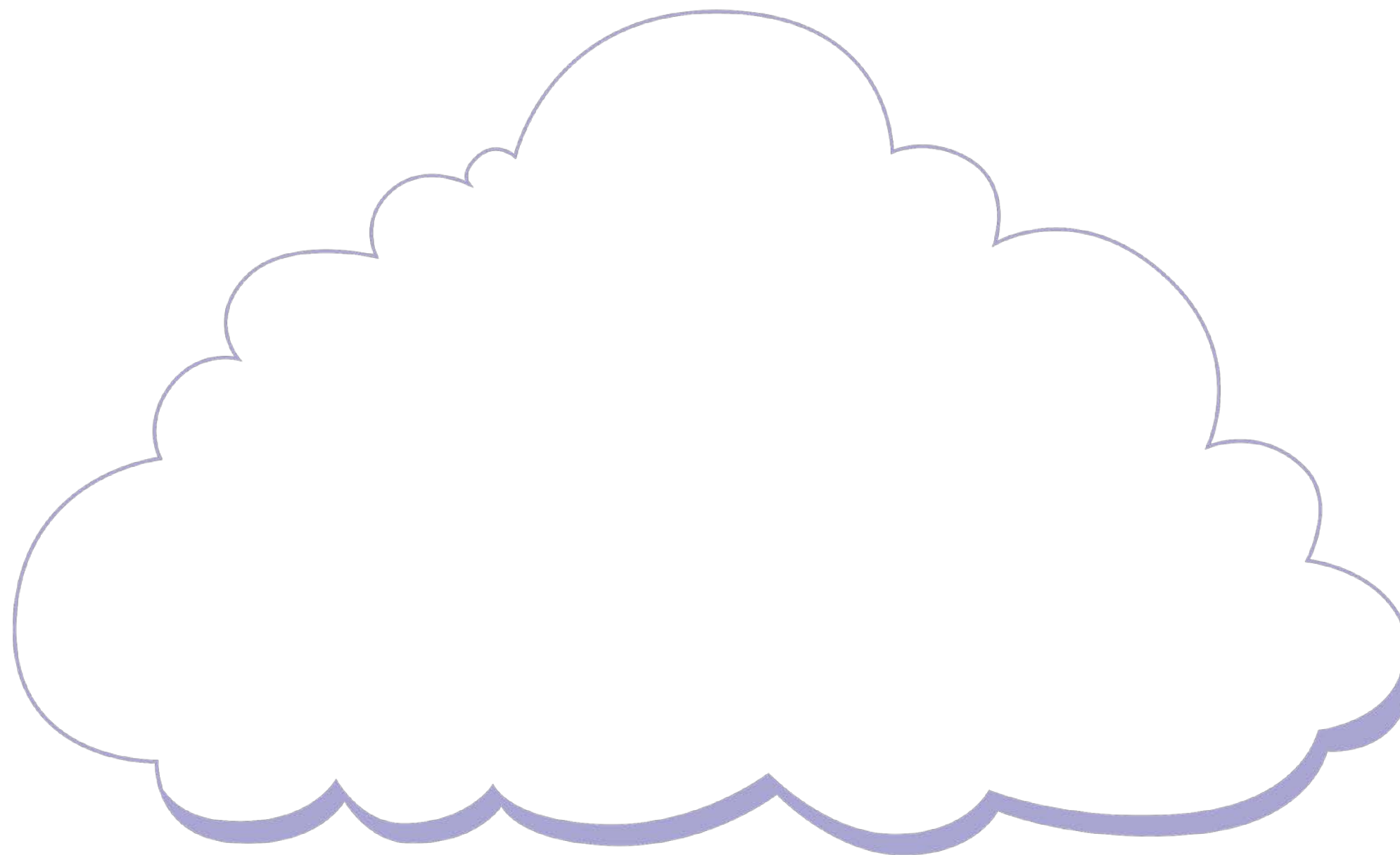
- Como vocês fariam a decomposição de problema para este caso?

Nuvem de palavras



© Getty Images

O que nós
**aprendemos
hoje?**





© Getty Images

O que nós
**aprendemos
hoje?**

Então ficamos assim...

- 1** A decomposição de problemas em Ciência de Dados divide um grande problema em partes menores para análise mais fácil e eficiente, ajudando a entender melhor cada parte e suas interações;
- 2** Na modelagem, cada subproblema é analisado com técnicas específicas, usando dados e hipóteses para criar modelos que ajudam na análise e decisão;
- 3** Ambas as etapas fazem parte do escopo de decomposição de problemas em Ciência de Dados.

Saiba mais

Aproveite esse curso gratuito da Alura para saber mais sobre problemas de classificação, como o problema de classificação de reservas de hotéis.

ALURA. Curso de Classificação: selecionando features. Disponível em:
<https://www.alura.com.br/conteudo/classificacao-selecionando-features>.

Acesso em: 10 maio 2024.

Referências da aula

CAMARGO, R; RIBAS, T. *Gestão ágil de projetos*: as melhores soluções para suas necessidades. São Paulo: Saraiva, 2019.

CHIAVENATO, I. *Administração*: teoria, processo e prática. São Paulo: Manole, 2014.

Identidade visual: Imagens © Getty Images

**Educação
Profissional
Paulista**

Técnico em
**Ciência de
Dados**