



Mineração de Dados

Agente Educacional

Sérgio M. Dias

Agenda

Knowledge discovery in databases

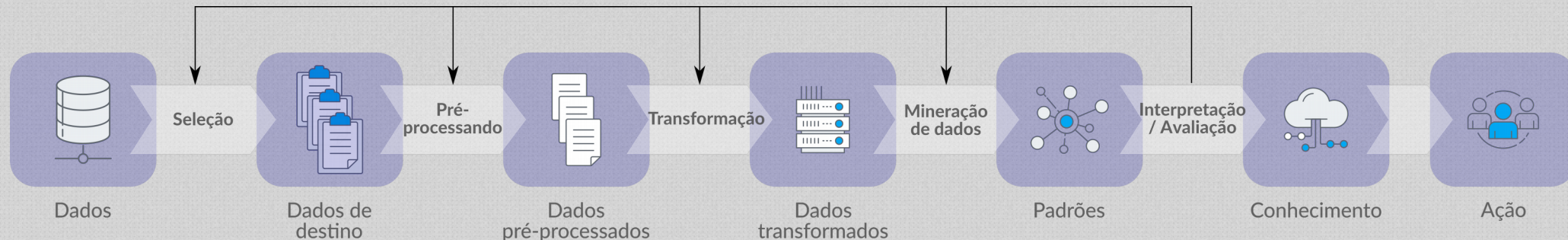
Mineração de Dados

Detecção de padrões

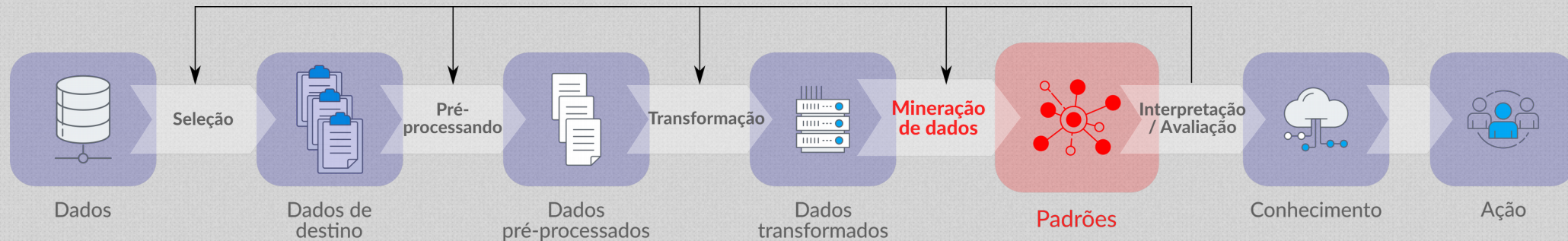


Knowledge Discovery in Databases (KDD)

Knowledge discovery in databases (KDD) é o processo de **extração não trivial de conhecimento** previamente desconhecido e potencialmente útil de bases de dados.



Importante etapa no fluxo de KDD que consiste na aplicação de técnicas computacionais para descoberta de **padrões relevantes** em grandes conjuntos de dados.



Mineração de Dados (Data Mining)

MD envolve obrigatoriamente:

- Conhecimento do negócio
- Técnicas adequadas para cada tipo de problema
- Infraestrutura de TIC (hardware, software, rede).

...

Constitui um processo bem mais próximo da pesquisa e investigação do que do desenvolvimento de software.

Em cenários de grandes volumes de dados, soluções adotadas devem atender a requisitos rígidos de escalabilidade.

Nem tudo se resolve com mais recurso de hardware!

Processo de exploração de dados à procura de padrões consistentes para detectar relacionamentos sistemáticos entre variáveis.

Usada para responder perguntas não triviais, do tipo:

- **Por que** determinados clientes são mais rentáveis?
- **Qual** tipo de cliente terá interesse pelo meu novo produto?
- **Quais** clientes podem ir para meus competidores?
- **Quais** os hábitos de meus clientes?

Baseada em modelos:

- Abstração matemática que representa um comportamento
- Normalmente uma função, $y = f(x)$

Mas o que são padrões?

Observe atentamente a sequência de letras abaixo e tente encontrar alguma característica marcante.

ABCXYABCZKABDKCABCTUABEWLABCWO

Passo 1: a primeira etapa é perceber que **existe uma sequência de letras** que se repete bastante. Encontramos as sequências "AB" e "ABC" e observamos que elas **ocorrem com frequência superior** à das outras sequências.

ABCXY**ABC**ZK**AB**DK**ABC**TU**AB**EW**ABC**WO

Passo 2: após determinarmos as sequências "ABC" e "AB", verificamos que elas **segmentam o padrão original em diversas unidades** independentes.

ABCXY ABCZK ABDKC ABCTU ABEWL ABCWO

Mineração de Dados

Mas o que são padrões?

Passo 3: fazem-se agora induções, que geram algumas representações genéricas dessas unidades.

ABC?? ABD?? ABE?? AB???

Suponha que a sequência **represente o histórico de compras de um cliente** em um supermercado, e que a ocorrência de:

A significa “**pão**”

B significa “**leite**”

C indica que o leite é do tipo “**desnatado**”

...

Mineração de Dados

Mas o que são padrões?

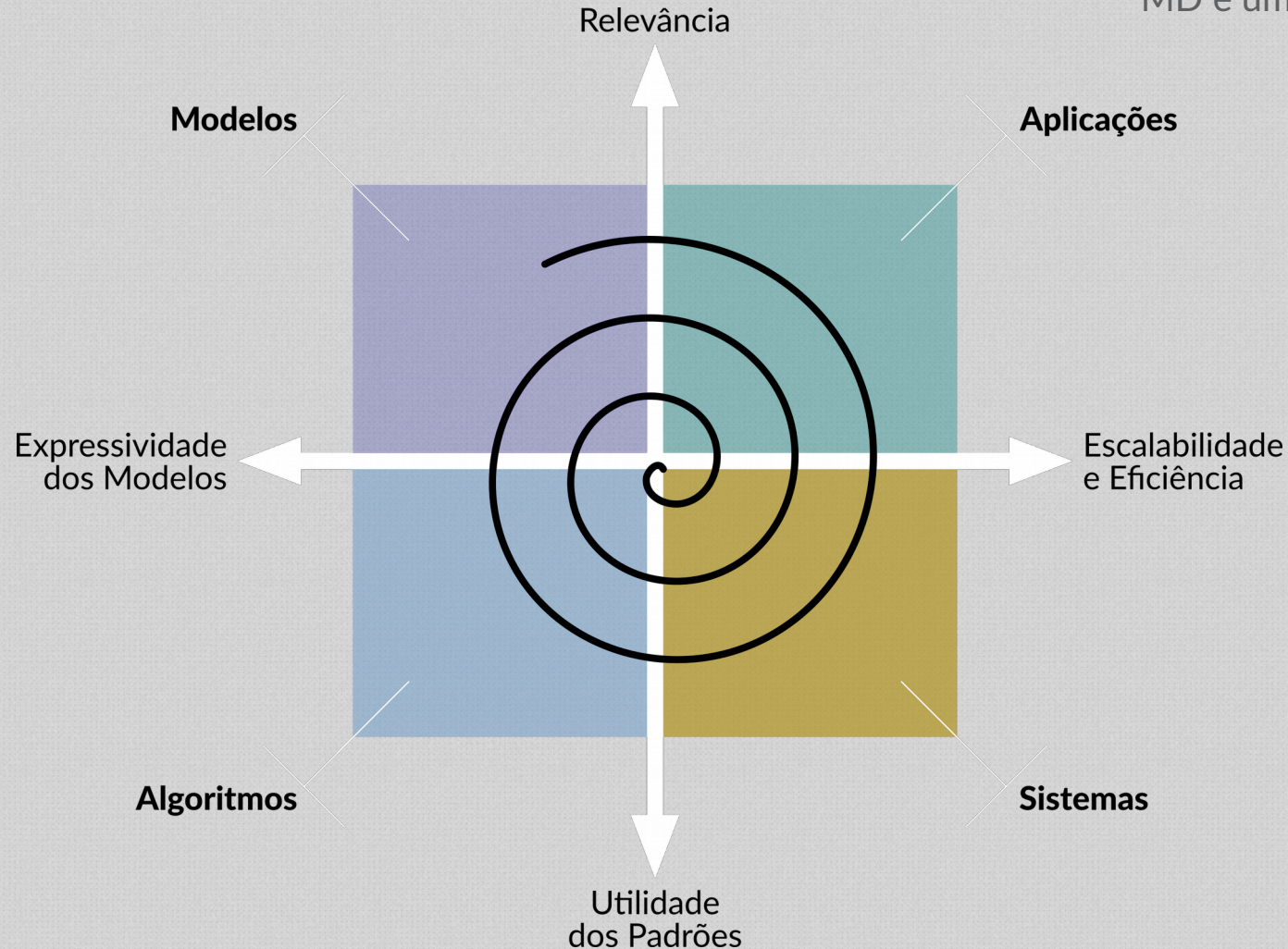
...

A descoberta de uma regra com as letras **AB** significa que toda vez que este cliente compra pão, também compra leite.

Os atributos foram associados por um processo de **reconhecimento de padrões**.

Mineração de Dados

MD é um Processo Iterativo



Obrigado!

Agente Educacional

Sérgio M. Dias

sergio.dias@serpro.gov.br | #31 6539

Demais agentes educacionais sobre o assunto

Marcelo Pita | marcelo.pita@serpro.gov.br | #81 8794

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950