**Mineração de Dados (Data Mining)**

* Conceito: geramos dados o tempo todo, no rastro que deixamos ao fazer compras de produtos ou serviços, nas plataformas sociais, trocas de mensagens, etc. ​Humanos também passaram a ter sede de dados.  A simples manutenção de um blog já gera dados interessantes. Grandes empresas sabem muito a nosso respeito, ainda que muitas vezes não tenhamos consciência disso.
* Análise de dados: dados são fatos brutos que representam eventos que ocorrem nas organizações ou no ambiente físico, antes de  serem organizados e arranjados de uma forma que as pessoas possam entendê-los e usá-los. Dados precisam ser lapidados para que se tornem informações úteis.

*Por que queremos dados?*

* Por que dados aparentemente sem muito sentido podem resultar em informações valiosas. Se dispomos de sistemas computacionais adequados, a seguinte transformação pode ocorrer:
* A ideia de big data vem exatamente daí, da necessidade de extração de conhecimento a partir de muitos dados.
* Junto ao big data está a ideia de mineração de dados (data-mining).
* Data-mining ou mineração de dados é o processo de explorar grandes quantidades de dados à procura de padrões para detectar novos relacionamentos entre variáveis.
* Procura descobrir padrões, tendências e correlações ocultas nos dados.
* Geralmente ainda está identificada com algoritmos.
* Mineração de dados é usada para :
* *Explicar: … Por que caiu a venda de sorvetes no Rio de Janeiro?*
* *Confirmar uma hipótese: uma companhia de seguros pode querer examinar os registros de seus clientes para determinar se famílias de duas rendas têm mais probabilidade de adquirir um plano de saúde do que famílias de uma renda.*
* *Explorar: analisar os dados buscando relacionamentos novos e não previstos. Uma companhia de cartão de crédito pode analisar seus registros históricos para determinar que fatores estão associados a pessoas que representam riscos para créditos.*

* O que é BI: baseia-se na capacidade de disseminar informações de forma rápida e segura auxiliando em processos de tomada de decisões. As organizações recolhem informações a fim de avaliar o ambiente de negócio e converter em campos significativos, tais como pesquisa de mercado, de indústria, de marketing e análise de competidores. Utiliza técnicas de recuperação da informação, inteligência artificial, reconhecimento de padrões, estatísticas. DM é usual em grandes bancos de dados e o resultado final pode ser exibido por meio de regras, hipóteses, árvores de decisão, etc. DM evolui como evolui o conhecimento e a inteligência empresarial.

Aplicações do data-mining:

1. Lucratividade
2. Retenção de clientes
3. Segmentação de clientes
4. Propensão de clientes
5. Otimização de canais
6. Marketing dirigido
7. Gerenciamento de risco
8. Prevenção de fraudes
9. Análise de cesta de compras
10. Previsao de demanda
11. Otimização de preços

* Wal- Mart: a cadeia de lojas  Wal-Mart, que identificou um hábito curioso dos consumidores: *Ao procurar eventuais relações entre o volume de vendas e os dias da semana, o software de data mining apontou que, às sextas-feiras, as vendas de cervejas cresciam na mesma proporção que as de fraldas. Ao comprar fraldas para seus bebês, os pais aproveitavam para abastecer o estoque de cerveja para o final de semana.*

Data mining opera de acordo com diferentes métodos:

1. Método de visualização
2. Método estatístico
3. Método de dedução
4. Método de indução
5. Método de estruturas de decisão (árvores)
6. Método de Redes Neurais
7. Método de Associação
8. Método da Cesta de Compras
9. Método de algoritmos genéticos

* Método de indução de regras: conjunto de regras que classificam os conjuntos de dados

- Ex: Se Renda > 60.000 e Débito < 10%, então Risco “bom” =>  Aprovar!

**Sistemas de apoio à decisão / modelagem analítica**

* Análise de Sensibilidade

Observar como mudanças repetidas em uma única variável afetam outras variáveis.

*Exemplo:* *reduzir a propaganda em 1000 reais repetidamente de forma que possamos entender sua relação com as vendas.*

* Análise de Otimização

Encontrar um valor ótimo para variáveis selecionadas, dadas certas restrições.

*Exemplo:* *qual o melhor montante de propaganda considerando nosso orçamento e escolha de mídia?*

* Técnica de cesta de compras para gerar regras de associação

As regras de associação são bastante utilizadas em banco de dados de tamanho grande e o resultado depende do algoritmo usado.

O nome “análise da cesta de compras” tem origem no algoritmo que começou classificando o tipo de cliente de um supermercado de acordo com a composição da sua cesta de compras.

* Técnica de cesta de compras para gerar regras de associação (cont.)

Associação sequencial:

Descobre associações que ligam eventos ao longo do tempo (ou seja, identifica padrões sequenciais).

*Exemplo:* *clientes que abrem uma conta corrente e após três meses abrem uma conta poupança e abrirão uma conta de cartão de crédito dentro de seis meses em 24% dos casos.*

Método de associação: relações significativas entre itens de dados armazenados. O objetivo é encontrar tendências a partir de um grande número de transações.

*Exemplo:*varrer registros de terminais de pontos de venda e descobrir que itens são vendidos juntos para redefinir disposição a campanhas.

*Exemplo:*Quando são comprados salgadinhos de milho, em 55% dos casos é comprado um refrigerante, como a coca-cola, a menos que haja uma promoção, caso em que a coca-cola é comprada em 75% durante a promoção.

Importante: é muito provável que o modelo inicial não atenda os objetivos do exercício de mineração de dados, sendo necessárias muitas repetições, especialmente entre as fases de projeto e de análise de dados. Isso envolve tentativas de diferentes técnicas de mineração de dados ou parâmetros em diferentes subconjuntos de dados antes de chegar a um resultado bem-sucedido!

Desafios: os softwares são trabalhosos e frustram quem espera respostas rápidas. Não há mágica. Os softwares de DM estão muito longe de descobrirem conhecimento sozinhos.

​

*Como selecionar a ferramenta de data-mining?*

1. 1 - Que algoritmos suporta?
2. 2 - Que saídas gera?
3. 3 - Que formatação de entrada exige?
4. 4 - Como a ferramenta/serviço adquire os dados para sua utilização?
5. 5 – Como o analista interage com a ferramenta?

É gráfica? / É linha de comando?​

1. 6 – Que nível de experiência é exigido desse profissional? Que pré-requisitos necessita?
2. 7 – A ferramenta suporta valores contínuos ou discretos?
3. 8 – Qual o limite de carga?
4. 9 – É focada em problemas específicos?
5. 10 – É uma ferramenta que “aprende”?

Ex: possui um modelo de aprendizagem interno?​

1. 11 – Que organizações estão utilizando essa ferramenta? É exclusiva de um modelo de empresa?
2. 12 – Que resultado já foram obtidos por essa ferramenta? Qual a relação de custo/benefício projetada?

     Escolha da ferramenta:

* **Acesso a fonte de dados heterogêneas**
* **Integração de conjuntos de dados**
* **Facilidade de incluir novas operações**
* **Facilidade para incluir novos métodos**
* **Recursos para planejamento de ações**
* **Processamento paralelo/distribuído**
* **Métodos disponíveis**

- No setor público é possível fazer o cruzamento de dados entre o estado civil de um funcionário e o salário que ele ganha, para verificar se isso tem influência na sua vida pessoal.

- Empresas como cadeias de supermercados podem recorrer a cruzamentos de dados para determinar que produtos são comprados em conjunto. Se um cliente X também compra o produto Y, talvez seja uma boa ideia posicionar os dos produtos perto, para facilitar a compra por parte do cliente.

* Ferramentas altamente especializadas.
* Elevado custo, em geral, pouco amigável.
* Voltada para grandes volumes de dados.
* Não há ferramentas melhores do que outras, depende da aplicação e interesses.
* O fator humano é preponderante.

​

Resultado:escolhas de softwares/ferramentas são quase empíricas, demandam elevada ação humana.