



Analytics e Business Intelligence

Interação 6

Prof. Dr. Felipe Aragão Campos Salles

Tema da Aula

Mineração de Dados



Definição de Mineração de Dados

“Data mining refere-se ao desenvolvimento do business intelligence a partir dos dados coletados, organizados e processados por uma empresa.

Técnicas de data mining estão sendo usadas por empresas para obter um melhor entendimento dos seus clientes e de suas próprias operações.

Neste capítulo, estudaremos algumas das principais técnicas de data mining, como por exemplo, árvores de decisão e análise de clusters, suas aplicações e os processos de data mining” (TURBAN, 2010).

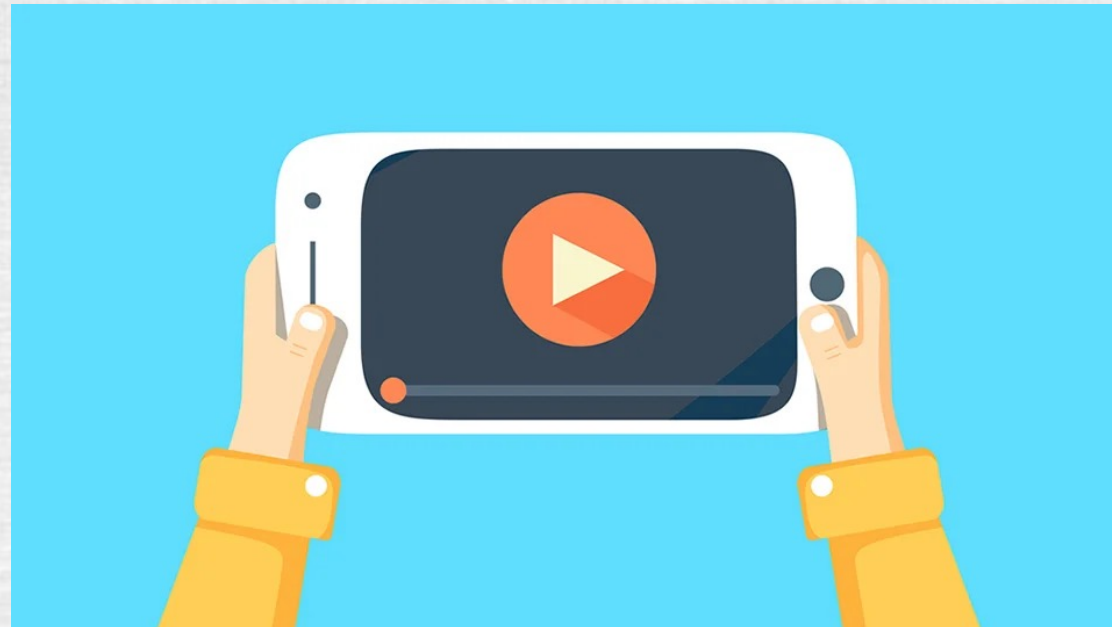
Exemplo 1:

Vinheta de Abertura – 4.1 (Turban, 2010)



Vídeo 1: Data Mining (Tecnologia em 3 minutos)

<https://www.youtube.com/watch?v=dWL0jvj1PJw>



Vídeo 2: Entenda porque a mineração de dados é tão importante

<https://www.youtube.com/watch?v=hLPwgfZ40pA>



Algumas áreas de aplicação do Data Mining

- Marketing;
- Bancos;
- Varejo e vendas;
- Fabricação e produção;
- Seguros;
- Empresas aéreas;
- Etc.

Alguns pontos importantes sobre Data Mining

- Muitas vezes, os dados estão escondidos em bancos de dados muito grandes que, às vezes, contêm dados de diversos anos. Em muitos casos, os dados estão limpos e consolidados em um *data warehouse*.
- O ambiente de *data mining* geralmente é uma arquitetura cliente/servidor ou uma arquitetura baseada na Web.
- Novas ferramentas sofisticadas, incluindo ferramentas de visualização avançada, ajudam a remover informações escondidas em arquivos corporativos ou arquivadas em registros públicos. Descobri-las envolve mexer e sincronizar os dados para a obtenção dos resultados certos. *Data miners* modernos também estão explorando a utilidade dos dados temporários (ou seja, texto não estruturado armazenado em lugares como bancos de dados Lotus Notes, arquivos de texto na Internet ou nas *intranets* de toda a empresa).

- Muitas vezes, o *miner* é um usuário final, capacitado por *datadrills* e outras ferramentas poderosas de consulta para fazer perguntas *ad hoc* e obter respostas rapidamente, requerendo pouca ou nenhuma habilidade de programação.
- Tirar a sorte grande muitas vezes envolve descobrir um resultado inesperado e exige que os usuários finais pensem de forma criativa.
- Ferramentas de *data mining* são facilmente combinadas com planilhas e outras ferramentas para desenvolvimento de *software*. Dessa forma, os dados extraídos podem ser analisados e processados de maneira rápida e fácil.
- Devido às grandes quantidades de dados e iniciativas sólidas de pesquisa, às vezes é necessário usar processamento paralelo para *data mining*.

Fonte: Turban (2010)

Métodos para a Identificação de Padrões em Dados

- Modelos simples (p.ex., consultas baseadas em SQL, processamento analítico *online* [OLAP], raciocínio humano)
- Modelos intermediários (p.ex., regressão, árvores de decisão, agrupamento)
- Modelos complexos (p.ex., redes neurais, outra indução de regras)



Exemplo 2:

Caso de Aplicação 4.1 (Turban, 2010)



Exemplo 3:

Caso de Aplicação 4.2 (Turban, 2010)



Exemplo 4:

Caso de Aplicação 4.3 (Turban, 2010)



Deep Learning

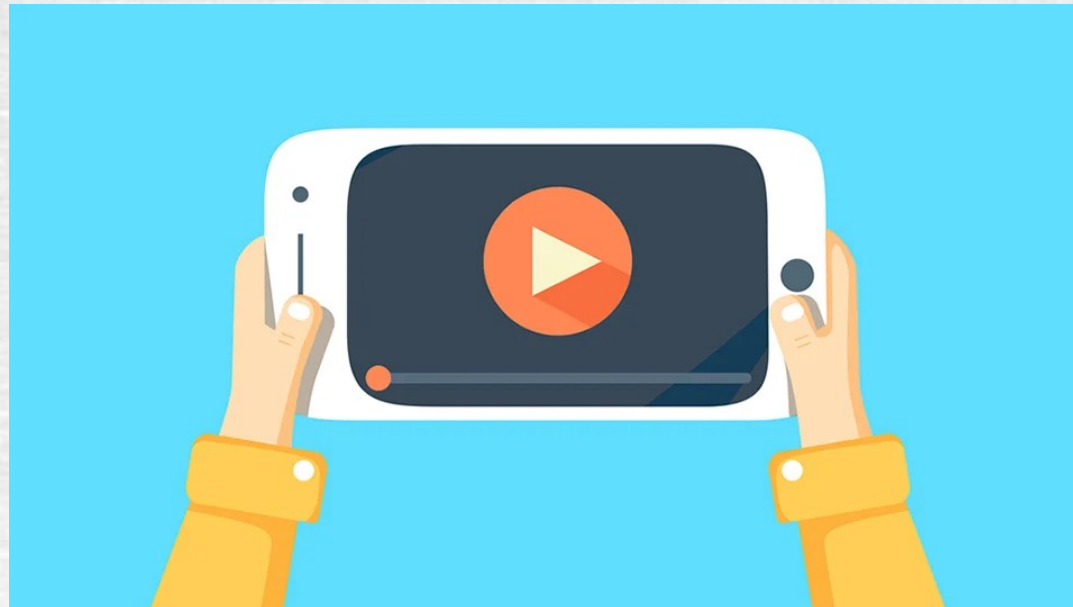
“O crescimento da era digital levou a um salto colossal nos dados gerados pelo usuário médio. Estes dados crescentes têm diversas aplicações: as empresas podem utilizá-los para dar um toque mais personalizado aos seus serviços, os governos podem utilizá-los para alocar melhor os seus fundos e as empresas podem utilizá-los para selecionar os melhores candidatos para um emprego. Embora estas aplicações possam parecer extremamente atraentes, há alguns problemas que devem ser resolvidos primeiro, nomeadamente, a recolha de dados e a extração de padrões úteis a partir dos dados” (SHARMA; TOKAS; ADLAKHA, 2021, tradução nossa).

“As disciplinas de mineração de dados e big data lidam com esses problemas, respectivamente. Mas, como já discutimos, a quantidade de dados é tão vasta que qualquer abordagem manual é extremamente demorada e dispendiosa. Portanto, isso limita os resultados potenciais desses dados. Este problema foi resolvido com a aplicação de aprendizagem profunda” (SHARMA; TOKAS; ADLAKHA, 2021, tradução nossa).

Deep Learning

Vídeo 3: Deep Learning = Dicionário do Programador

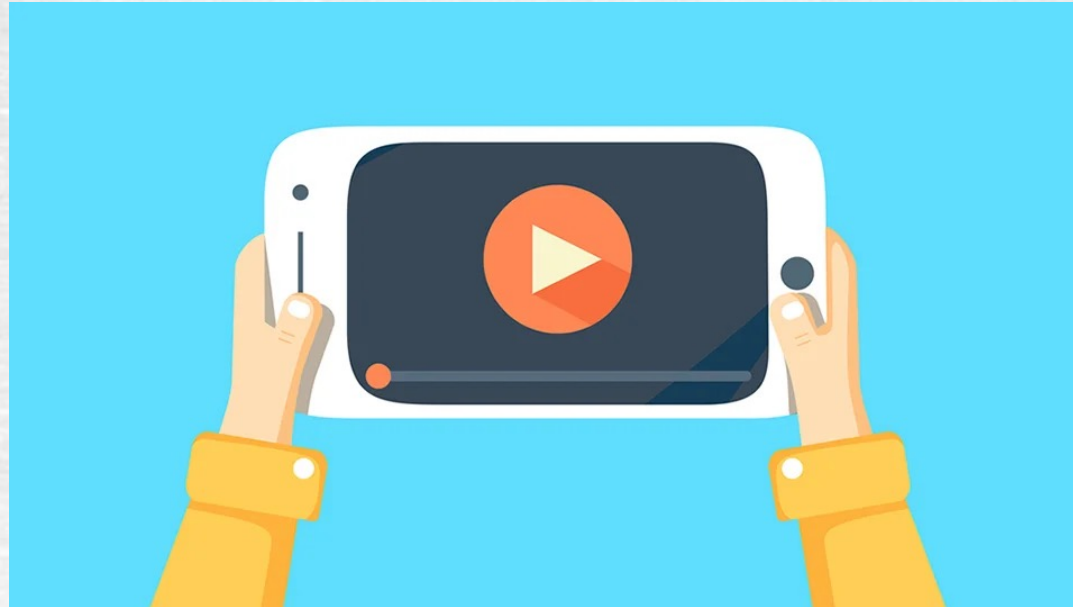
<https://www.youtube.com/watch?v=7EpQsB225vs>



Diferença entre Machine Learning e Deep Learning

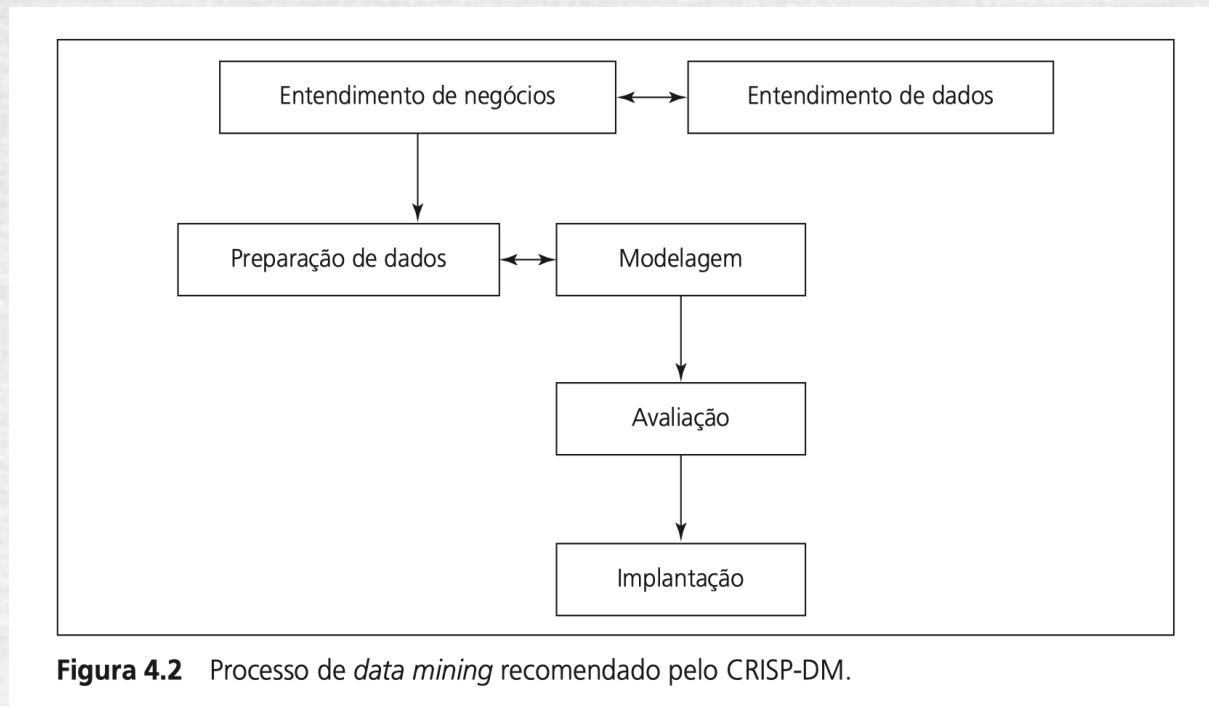
Vídeo 4:

<https://www.youtube.com/watch?v=q6kJ71tEYqM>



Metodologias para Implementação de Projetos de Data Mining

1) Cross-Industry Standard Process for Data Mining



2) Seis Sigma



Figura 4.3 Processo de *data mining* baseado no Six Sigma.



Erros crassos em *data mining*

Muitas vezes, os 10 erros de *data mining* a seguir são cometidos na prática, e deve-se tentar evitá-los:

1. Selecionar o problema errado para *data mining*.
2. Ignorar o que seu patrocinador pensa sobre o que é *data mining* e o que realmente pode e não pode fazer.
3. Não dedicar tempo suficiente para preparação dos dados. Exige mais esforço do que normalmente é deduzido.
4. Olhar apenas resultados reunidos e não os registros individuais. O IMS DB2 da IBM consegue destacar registros individuais de interesse.
5. Ser negligente quanto ao controle de procedimento e resultados do *data mining*.
6. Ignorar descobertas suspeitas e ir adiante rapidamente.
7. Executar algoritmos de mining de modo repetido e cego. É importante pensar bastante sobre o próximo estágio da análise de dados. O *data mining* é uma atividade com muita participação ativa.
8. Acreditar em tudo que se ouve sobre os dados.
9. Acreditar em tudo que se ouve sobre sua própria análise de *data mining*.
10. Medir seus resultados de maneira diferente daquela que seu patrocinador usa para medir os dele.

Fontes: adaptado de D. Skalak, “Data Mining Blunders Exposed!” *DB2 Magazine*, Vol.6, No. 2, Second Quarter 2001, pp. 10–13; and R. Shultz, *Live from NCDM: Tales of Database Buffoonery*, December 7, 2004, directmag.com/news/ncdm-12-07-04/index.html (acessado em abril de 2006).

[illegible]

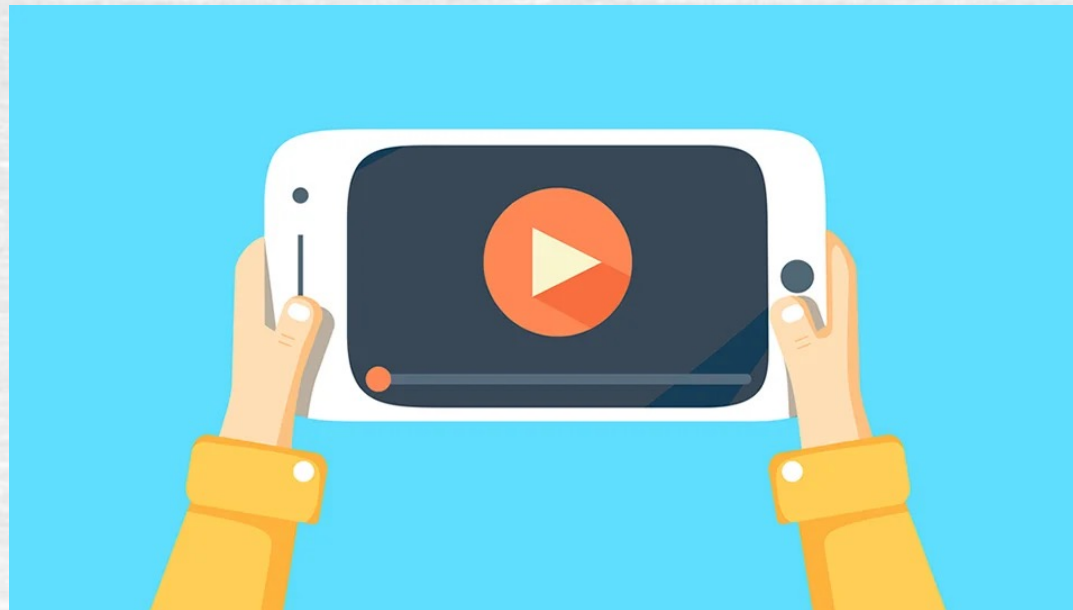
Text mining é a aplicação de data mining em arquivos de texto não estruturados ou menos estruturados. O data mining aproveita-se da infraestrutura de dados armazenados para extrair informações adicionais úteis (TURBAN, 2010).



Exemplo de Text Mining

Vídeo 5:

<https://www.youtube.com/watch?v=UqkkgTaF9RU>



Referências

TURBAN, E. **Business Intelligence: um enfoque gerencial**. Porto Alegre: Ed. Bookman, 2009.

SHARMA, D. K.; TOKAS, B.; ADLAKHA, L. Deepak Chapter 2 - Deep learning in big data and data mining. In: PIRUI *et al.* **In Hybrid Computational Intelligence for Pattern Analysis**. Trends in Deep Learning Methodologies, Academic Press, 2021, Pages 37-61. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/B9780128222263000027>>. Acesso em: 11/04/2024.

Obrigado!