MINERAÇÃO DE DADOS

Prof. Me. Napoleão Póvoa Ribeiro Filho

UNITINS

GOVERNO DO TOCANTINS
TRABALHANDO E CUIDANDO DE TODOS

MINERAÇÃO DE DADOS

- Durante as décadas de 1980 e 1990, com o aumento da capacidade de armazenamento e processamento dos computadores, os bancos de dados passaram a acumular enormes quantidades de informações.
- No entanto, transformar esses dados em conhecimento ainda era um desafio.

MINERAÇÃO DE DADOS

O termo surgiu nos anos 90 para descrever a união de estatística e ciência da computação (machine learning e bancos de dados), aplicada a grandes volumes de dados em ciência, engenharia e negócios.

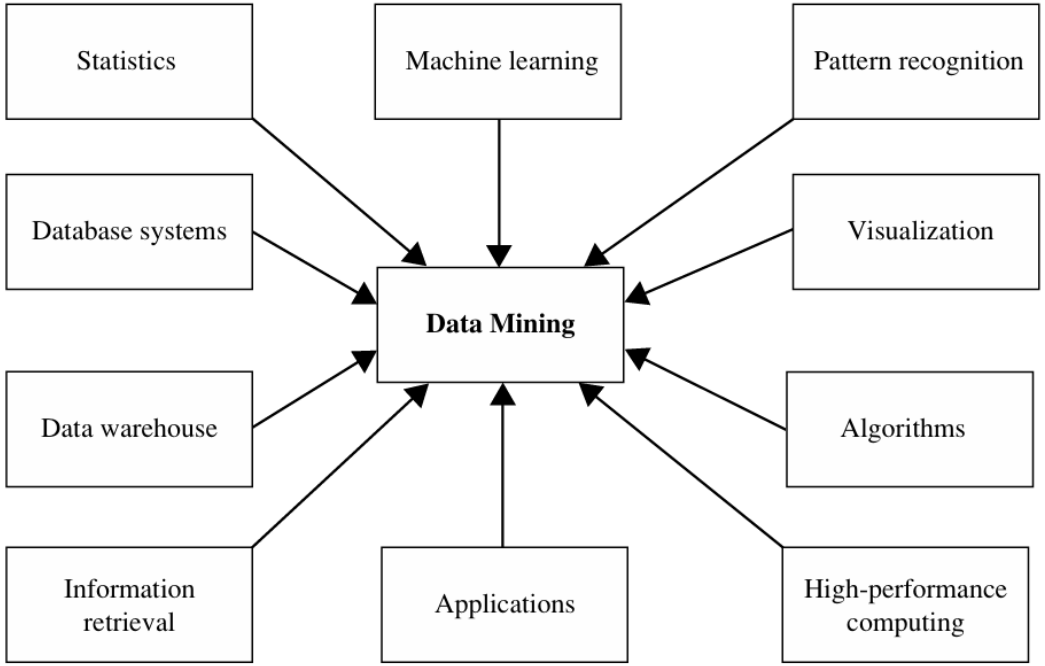
UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MINERAÇÃO DE DADOS

- “Processo de descoberta automática de informações úteis em grandes depósitos de dados” [Tan et. al., 2006].
- “Data mining is the application of specific algorithms for extracting patterns from data” [Fayyad et. al., 1996]
- O processo, não trivial, de extrair informação implícita, potencialmente útil e previamente desconhecida de dados.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

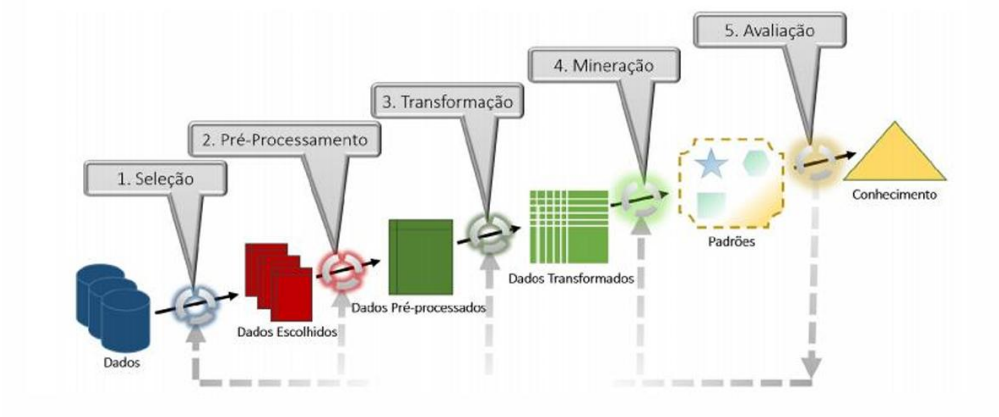
MINERAÇÃO DE DADOS



UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

KDD (*Knowledge Discovery from Data*)

- KDD é um processo que compreende os passos comuns desde a coleta de dados em um banco de dados até a obtenção de padrões úteis e previamente desconhecidos.



UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

ETAPAS DO KDD

- Seleção dos dados
- Pré-processamento dos dados
- Transformação dos dados
- Mineração dos dados
- Interpretação/Avaliação/Apresentação

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

SELEÇÃO DOS DADOS

- Nesta etapa, são escolhidos quais dados serão analisados, garantindo que sejam relevantes para o problema a ser resolvido.

- **Exemplo:** O e-commerce decide usar informações das compras realizadas nos últimos 12 meses, incluindo dados de clientes, produtos adquiridos, valores gastos e datas das transações.

PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

- Consiste em corrigir erros, remover valores inconsistentes e tratar dados ausentes para evitar distorções na análise.
- Exemplo: O e-commerce encontra clientes cadastrados mais de uma vez com e-mails diferentes e valores nulos na idade dos clientes. Ele remove duplicatas e preenche valores ausentes com a média de idade dos clientes ativos.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

- Os dados são organizados e convertidos para um formato adequado, podendo incluir normalização, agregação e criação de novas variáveis.

- Exemplo: A empresa cria uma nova variável chamada "Frequência de Compra" para classificar clientes em três grupos: compra frequente, esporádica ou apenas em promoções.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MINERAÇÃO DE DADOS

- São aplicados algoritmos para identificar padrões, tendências ou relações nos dados, utilizando técnicas como agrupamento, classificação ou regras de associação.
- Exemplo: O e-commerce usa um algoritmo de clusterização (agrupamento) e descobre que clientes que compram acima de R$ 500 por mês têm maior fidelidade, enquanto os que compram apenas em promoções raramente retornam.

INTERPRETAÇÃO E AVALIAÇÃO DOS RESULTADOS

- Os padrões encontrados são analisados e validados para garantir sua utilidade na tomada de decisões.

- Exemplo: A equipe percebe que clientes frequentes valorizam benefícios exclusivos e decide criar um programa de fidelidade, oferecendo descontos progressivos conforme o número de compras realizadas.

O QUE PODE SER MINERADO

- Diferentes tipos de padrões podem ser descobertos desde que:
  - Possua um volume da dados minimamente significativo
  - Seja passível de se questionar algo
  - Não seja trivial

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

O QUE PODE SER MINERADO

- Os principais tipos de conhecimento extraído incluem:
  - Padrões de associação
  - Padrões sequenciais
  - Classificação
  - Agrupamento (clustering)
  - Deteção de anomalias
  - Mineração de regras de decisão
  - Mineração de dados textuais e Web Mining

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

PADRÕES DE ASSOCIAÇÃO

- Identificam relações entre itens em um conjunto de dados, como regras do tipo "se X ocorre, então Y também ocorre frequentemente".

- **Exemplo:** Em um supermercado, descobre-se que clientes que compram pão também costumam comprar manteiga (Regra: *se comprar pão, há 80% de chance de comprar manteiga*).

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

PADRÕES SEQUENCIAIS

- Detectam padrões de eventos que ocorrem em sequência ao longo do tempo.

- Exemplo: Em um e-commerce, identifica-se que clientes que compram um smartphone geralmente compram um fone de ouvido algumas semanas depois.

CLASSIFICAÇÃO

- Atribui categorias a novos dados com base em padrões aprendidos de dados históricos.

- Exemplo: Um banco pode classificar clientes como "baixo risco" ou "alto risco" de inadimplência ao analisar histórico de crédito e comportamento de pagamento.

AGRUPAMENTO

- Agrupa dados semelhantes sem que categorias pré-definidas existam, revelando padrões ocultos.

- Exemplo: Um aplicativo de streaming descobre grupos de usuários com gostos musicais parecidos e sugere playlists personalizadas.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

DETECÇÃO DE ANOMALIAS

- Identifica eventos ou padrões que fogem do comportamento normal dos dados.

- Exemplo: Um sistema antifraude percebe que um cartão de crédito usado sempre no Brasil foi utilizado na China sem aviso prévio, possivelmente indicando fraude.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MINERAÇÃO DE REGRAS DE DECISÃO

- Extrai regras lógicas para tomada de decisão com base em variáveis do conjunto de dados.

- Exemplo: Um hospital pode extrair regras que ajudam a prever se um paciente tem alto risco de diabetes com base em idade, IMC e histórico familiar.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS

MINERAÇÃO DE DADOS TEXTUAIS

- Extrai conhecimento de textos e páginas da web para análise de sentimentos, recomendações ou segmentação de conteúdo.

- Exemplo: Empresas analisam avaliações de produtos para identificar sentimentos positivos ou negativos em comentários de clientes.

UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS


UNITINS
UNIVERSIDADE ESTADUAL DO TOCANTINS