



By @kakashi_copiador

RESUMO

DEFINIÇÕES DE DATA MINING

Data Mining é o processo de explorar grande quantidade de dados para extração não-trivial de informação implícita desconhecida.

Palavras-chave: exploração; informação implícita desconhecida.

Data Mining é uso de teorias, métodos, processos e tecnologias para organizar uma grande quantidade de dados brutos para identificar padrões de comportamentos em determinados públicos.

Palavras-chave: teorias; métodos; processos; tecnologias; organizar dados brutos; padrões de comportamentos.

Data Mining é a categoria de ferramentas de análise denominada open-end e que permite ao usuário avaliar tendências e padrões não conhecidos entre os dados.

Palavras-chave: ferramenta de análise; open-end; tendências e padrões.

Data Mining é o processo de descoberta de novas correlações, padrões e tendências entre as informações de uma empresa, por meio da análise de grandes quantidades de dados armazenados em bancos de dados usando técnicas de reconhecimento de padrões, estatísticas e matemáticas.

Palavras-chave: descoberta; correlações; padrões; tendências; reconhecimento de padrões; estatística; matemática.

Data Mining constitui em uma técnica para a exploração e análise de dados, visando descobrir padrões e regras, a princípio ocultos, importantes à aplicação.

Palavras-chave: exploração e análise de dados; padrões; regras; ocultos.

Data Mining é o conjunto de ferramentas que permitem ao usuário avaliar tendências e padrões não conhecidos entre os dados. Esses tipos de ferramentas podem utilizar técnicas avançadas de computação como redes neurais, algoritmos genéticos e lógica nebulosa (fuzzy), dentre outras.

Palavras-chave: tendências; padrões; redes neurais; algoritmos genéticos; lógica nebulosa.

Data Mining é o conjunto de ferramentas e técnicas de mineração de dados que têm por objetivo buscar a classificação e o agrupamento (clusterização) de dados, bem como identificar padrões.

Palavras-chave: classificação; agrupamento; clusterização; padrões.

Data Mining é o processo de explorar grandes quantidades de dados à procura de padrões consistentes com o intuito de detectar relacionamentos sistemáticos entre variáveis e novos subconjuntos de dados.

Palavras-chave: padrões; relacionamentos.

Data Mining consiste em explorar um conjunto de dados visando a extrair ou a ajudar a evidenciar padrões, como regras de associação ou sequências temporais, para detectar relacionamentos entre estes.

Palavras-chave: exploração; padrões; regras; associação; sequência temporal; detecção.

Data Mining são ferramentas que utilizam diversas técnicas de natureza estatística, como a análise de conglomerados (cluster analysis), que tem como objetivo agrupar, em diferentes conjuntos de dados, os elementos identificados como semelhantes entre si, com base nas características analisadas.

Palavras-chave: estatística; análise de conglomerados; agrupamento.

Data Mining é o conjunto de técnicas que, envolvendo métodos matemáticos e estatísticos, algoritmos e princípios de inteligência artificial, tem o objetivo de descobrir relacionamentos significativos entre dados armazenados em repositórios de grandes volumes e concluir sobre padrões de comportamento de clientes de uma organização.

Palavras-chave: métodos matemáticos e estatístico; inteligência artificial; relacionamentos; padrões; comportamentos.

Data Mining é o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados.

Palavras-chave: padrões; regras de associação; sequências temporais; relacionamentos.

Data Mining é o processo de identificar, em dados, padrões válidos, novos, potencialmente úteis e, ao final, compreensíveis.

Palavras-chave: padrões; utilidade.

Data Mining é um método computacional que permite extrair informações a partir de grande quantidade de dados.

Palavras-chave: extração.

Data Mining é o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais.

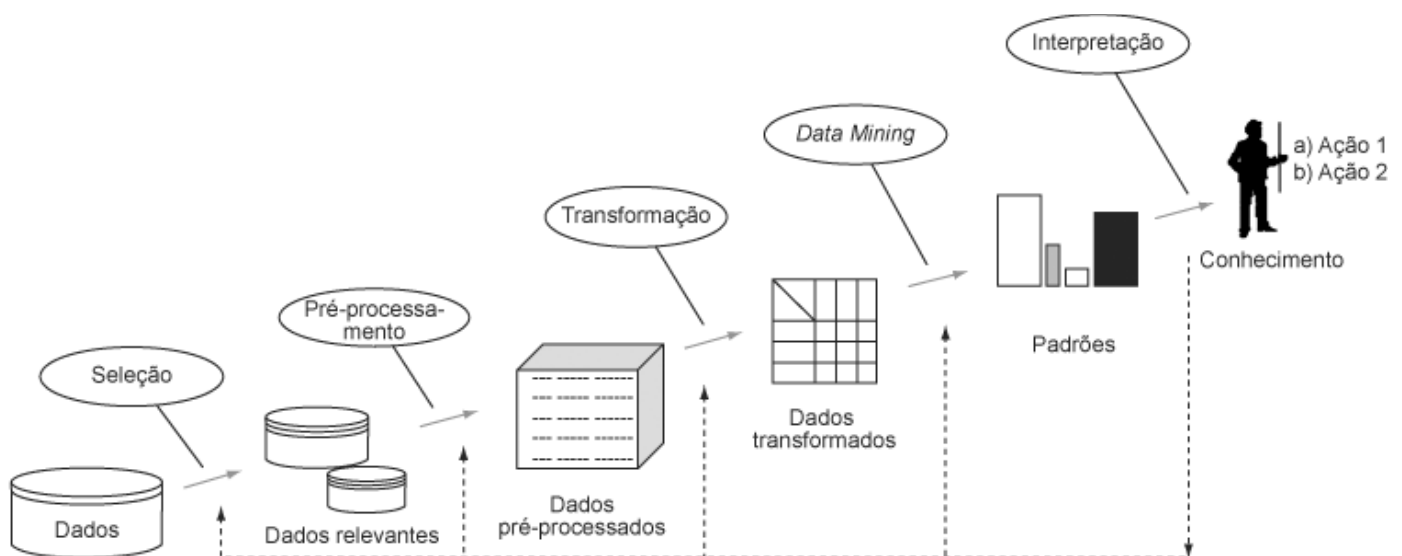
Palavras-chave: exploração; padrões consistentes; regras de associação; sequência temporal.

Data Mining é o processo de analisar de maneira semi-automática grandes bancos de dados para encontrar padrões úteis.

Palavras-chave: padrões.

PROCESSO DE DESCOBERTA DE CONHECIMENTO

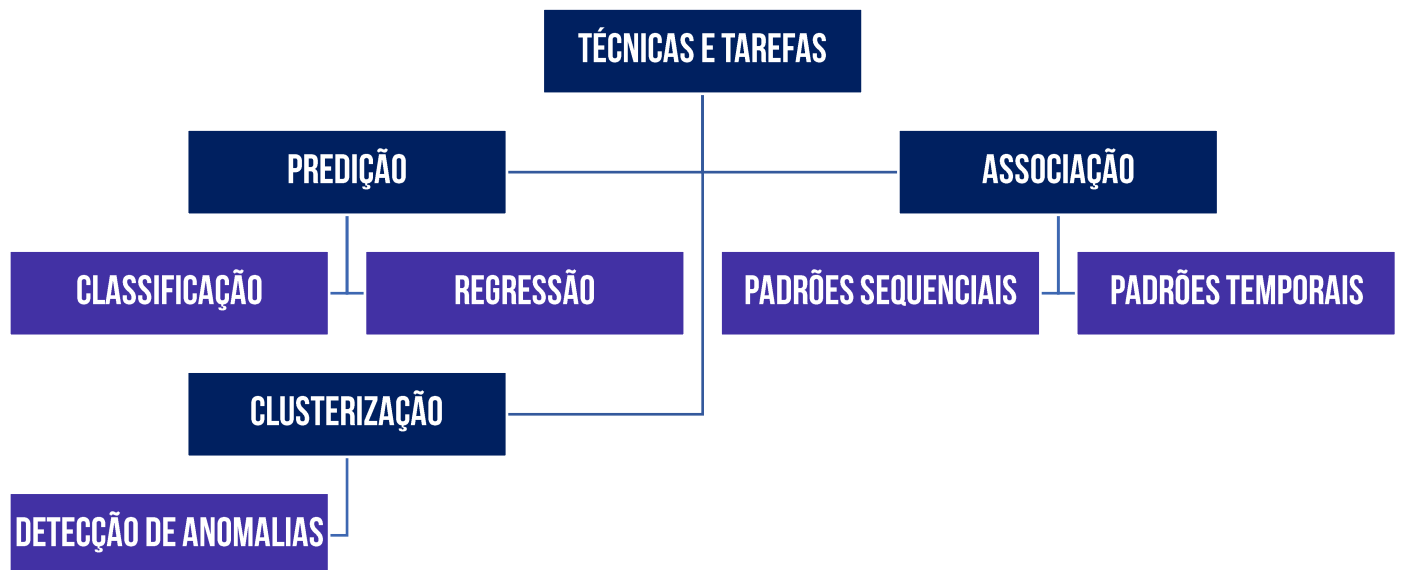
Data Mining faz parte de um processo muito maior de descoberta de conhecimento chamada KDD (Knowledge Discovery in Databases – Descoberta de Conhecimento em Bancos de Dados).





PREVISÃO IDENTIFICAÇÃO CLASSIFICAÇÃO OTIMIZAÇÃO

| OBJETIVOS | DESCRIÇÃO |
|---------------|--|
| PREVISÃO | A mineração de dados pode mostrar como certos atributos dos dados se comportarão no futuro. Para realizar a previsão (ou prognóstico), a lógica de negócios é utilizada em conjunto com a mineração de dados (Ex: prever um terremoto com alta probabilidade). |
| IDENTIFICAÇÃO | Padrões de dados podem ser usados para identificar a existência de um item, um evento ou uma atividade (Ex: padrões de comportamento de hackers permitem identificar possíveis intrusos acessando sistema). |
| CLASSIFICAÇÃO | A mineração de dados permite particionar os dados de modo que diferentes classes ou categorias possam ser identificadas com base em combinações de parâmetros (Ex: clientes podem ser categorizados pelos seus perfis de compradores). |
| OTIMIZAÇÃO | A mineração de dados pode otimizar o uso de recursos limitados, como tempo, espaço, dinheiro ou materiais e maximizar variáveis de saída como vendas ou lucros sob determinadas restrições (Ex: tempo, escopo e custo de um projeto). |



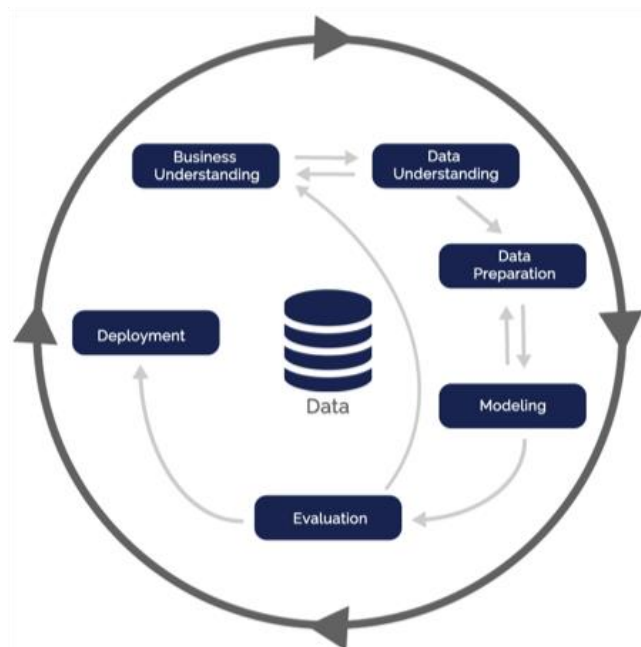
| TÉCNICAS | DESCRIÇÃO |
|---------------|---|
| CLASSIFICAÇÃO | Hierarquia de classes com base em um conjunto existente de eventos ou transações. |
| REGRESSÃO | Regra de classificação que é uma função sobre variáveis. |

| | |
|-----------------------------|---|
| | |
| REGRAS DE ASSOCIAÇÃO | Busca descobrir relacionamentos entre variáveis correlacionando a presença de um item com uma faixa de valores para outro conjunto de variáveis |
| AGRUPAMENTO | Particiona dados em segmentos previamente desconhecidos com características semelhantes |

| MEDIDAS DE INTERESSE | DESCRIÇÃO |
|---------------------------|--|
| SUORTE/PREVALÊNCIA | Trata-se da frequência com que um conjunto de itens específico ocorre no banco de dados, isto é, o percentual de transações que contém todos os itens em um conjunto (Ex: 50% das compras realizadas em um mercado contém arroz e refrigerante). |
| CONFIANÇA/FORÇA | Trata-se da probabilidade de que exista uma relação entre itens (Ex: 70% dos clientes que compram fraldas também compram cerveja). |

| CONCEITOS AVANÇADOS | DESCRIÇÃO |
|-------------------------------|--|
| APRENDIZADO DE MÁQUINA | Trata-se de uma ferramenta poderosa para a aquisição automática de conhecimento por meio da imitação do comportamento de aprendizagem humano com foco em aprender a reconhecer padrões complexos e tomar decisões. |
| MINERAÇÃO DE TEXTO | Trata-se de um meio para encontrar padrões interessantes/úteis em um contexto de informações textuais não estruturadas, combinado com alguma tecnologia de extração e de recuperação da informação, processo de linguagem natural e de sumarização ou indexação de documentos. |

| TÉCNICAS DE PRÉ-PROCESSAMENTO | DESCRIÇÃO |
|--|---|
| AGREGAÇÃO | Combina dois ou mais atributos (ou objetos) em um único atributo (ou objeto) com a finalidade de reduzir o número de atributos ou objetos, alterar escalas e tornar os dados mais estáveis. |
| AMOSTRAGEM | O princípio básico é: usar uma amostra funciona tão bem quanto usar o conjunto completo de dados, se a amostra for representativa. Ela é representativa se tiver aproximadamente as mesmas propriedades de interesse do conjunto inicial. |
| REDUÇÃO DE DIMENSIONALIDADE | Essa técnica reduz a quantidade de tempo e memória necessárias pelos algoritmos de mineração de dados, permitindo que os dados sejam mais facilmente visualizados e ajudando a eliminar características irrelevantes. |
| SELEÇÃO DE SUBCONJUNTOS DE RECURSOS | Trata-se de outra forma de reduzir a dimensionalidade dos dados, buscando eliminar características redundantes ou irrelevantes por meio de diversas abordagens diferentes. |
| CRIAÇÃO DE RECURSOS | Essa técnica busca criar novos atributos que podem capturar informação importante em um conjunto de dados muito mais eficientemente que os atributos originais. |
| BINARIZAÇÃO E DISCRETIZAÇÃO | Técnica que busca transformar dados para um formato de atributos binários ou discretos. |
| TRANSFORMAÇÃO DE VARIÁVEIS | Essa técnica busca melhorar a eficiência de algoritmos de classificação envolvendo redes neurais e auxiliar técnicas estatísticas que se baseiam na suposição da normalidade dos dados. |



| CRISP-DM | DESCRIÇÃO |
|-------------------------|--|
| ENTENDIMENTO DO NEGÓCIO | Busca compreender das necessidades gerenciais e dos objetivos e requisitos de negócio que devem ser atendidos pela mineração de dados. |
| ENTENDIMENTO DOS DADOS | Busca identificar os dados relevantes das diferentes fontes de dados. |
| PREPARAÇÃO DOS DADOS | Busca carregar os dados identificados no passo anterior e prepará-los para análise por métodos de mineração de dados. |
| CONSTRUÇÃO DO MODELO | Busca selecionar e aplicar técnicas de modelagem a um conjunto de dados previamente preparado. |
| TESTE E AVALIAÇÃO | Busca testar e avaliar os modelos desenvolvidos. |
| IMPLANTAÇÃO | Busca organizar o conhecimento adquirido com a exploração dos dados de forma que o usuário possa compreendê-lo. |