

MINERAÇÃO DE DADOS: UMA INTRODUÇÃO

Prof. Ronaldo R. Goldschmidt

ronaldo@de9.ime.eb.br

rribeiro@univercidade.br

geocities.yahoo.com.br/ronaldo_goldschmidt

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

6. CONCLUSÕES

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

6. CONCLUSÕES

1. INTRODUÇÃO E MOTIVAÇÃO

Fatos: Avanços em TI e o crescimento exponencial de BDs

- Internet
- SGBDs
- Dispositivos de memória (maior capacidade e menor custo)
- Leitoras de códigos de barras
- Sistemas de Informação em geral
- Etc...

1. INTRODUÇÃO E MOTIVAÇÃO

Exemplos:

- **FedEx**
- **UPS**
- **Wal-Mart**
- **NASA**
- **Projeto Genoma**

1. INTRODUÇÃO E MOTIVAÇÃO

Posicionamento:

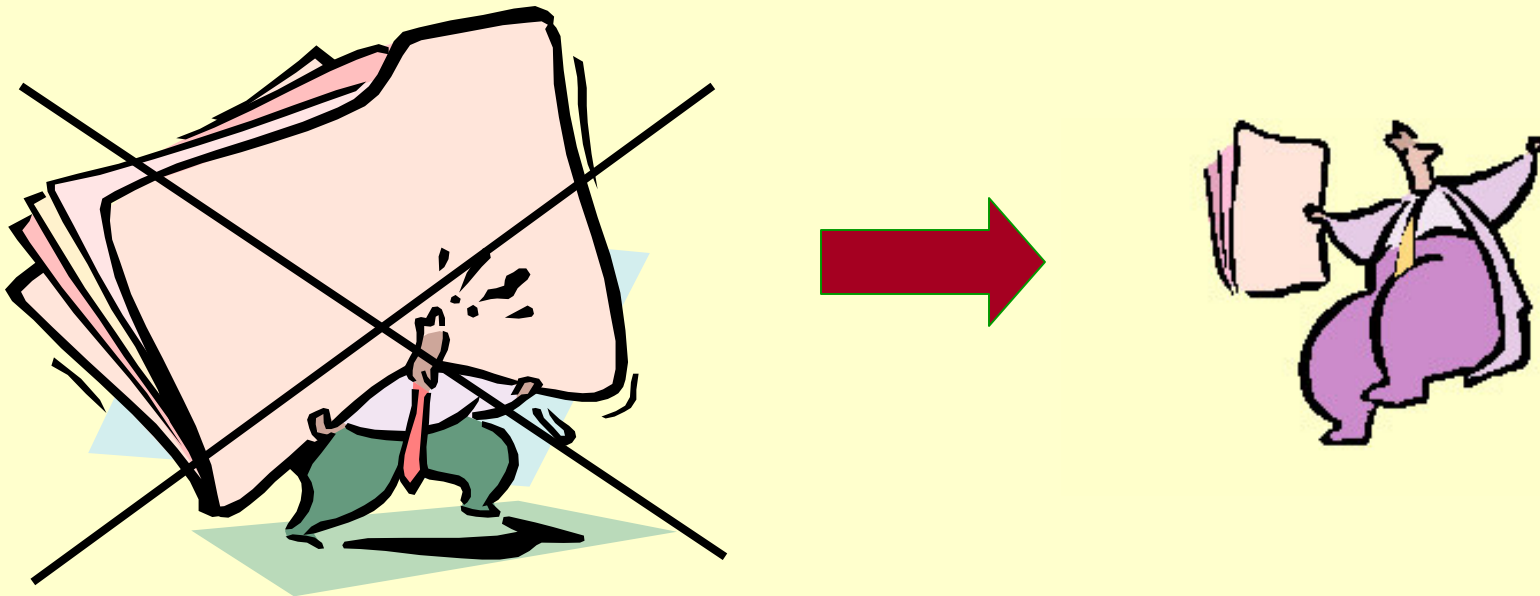
Hoje, a maioria das organizações produz **mais informações** em uma semana **do que muitas pessoas** poderiam ler em toda vida.

“Mais dados implicando em menos informações.”

1. INTRODUÇÃO E MOTIVAÇÃO

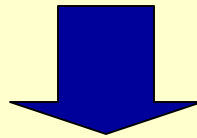
Necessidade:

Ferramentas **inteligentes** que auxiliem na **análise de dados** e na **busca por conhecimentos** em **GRANDES** bases de dados.

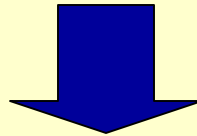


1. INTRODUÇÃO E MOTIVAÇÃO

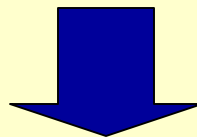
Avanços em TI



Crescimento Exponencial de BDs



Necessidade de Ferramentas para Análise Grandes BDs

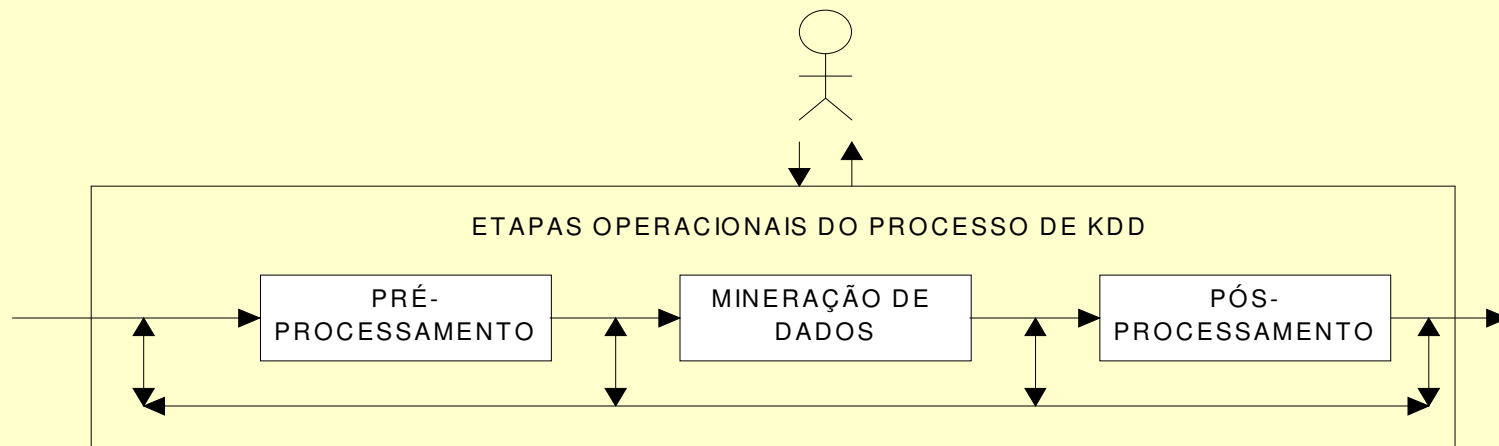


Área da Descoberta do Conhecimento em Bases de Dados (KDD)

1. INTRODUÇÃO E MOTIVAÇÃO

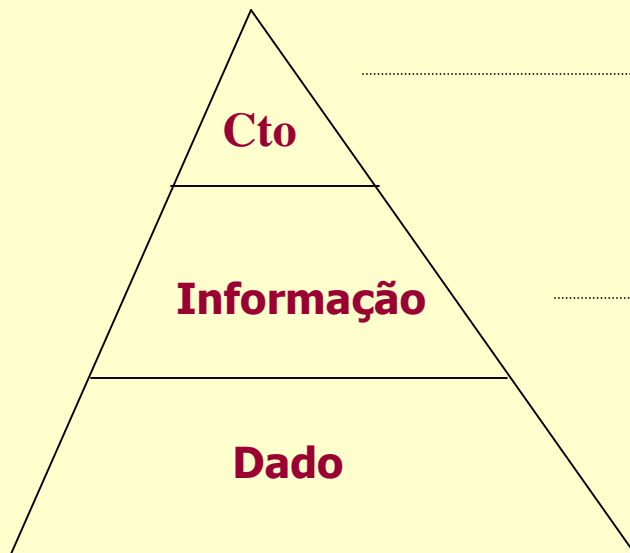
- **KDD – Knowledge Discovery in Databases**

“É um **processo**, de várias etapas, não trivial, **interativo e iterativo**, para **identificação** de **padrões compreensíveis, válidos, novos** e potencialmente **úteis** a partir de bases de dados.” [Fayyad et al., 1996]



1. INTRODUÇÃO E MOTIVAÇÃO

Hierarquia Dado - Informação - Conhecimento:



Se Capacidade Mensal de Endividamento > 60%

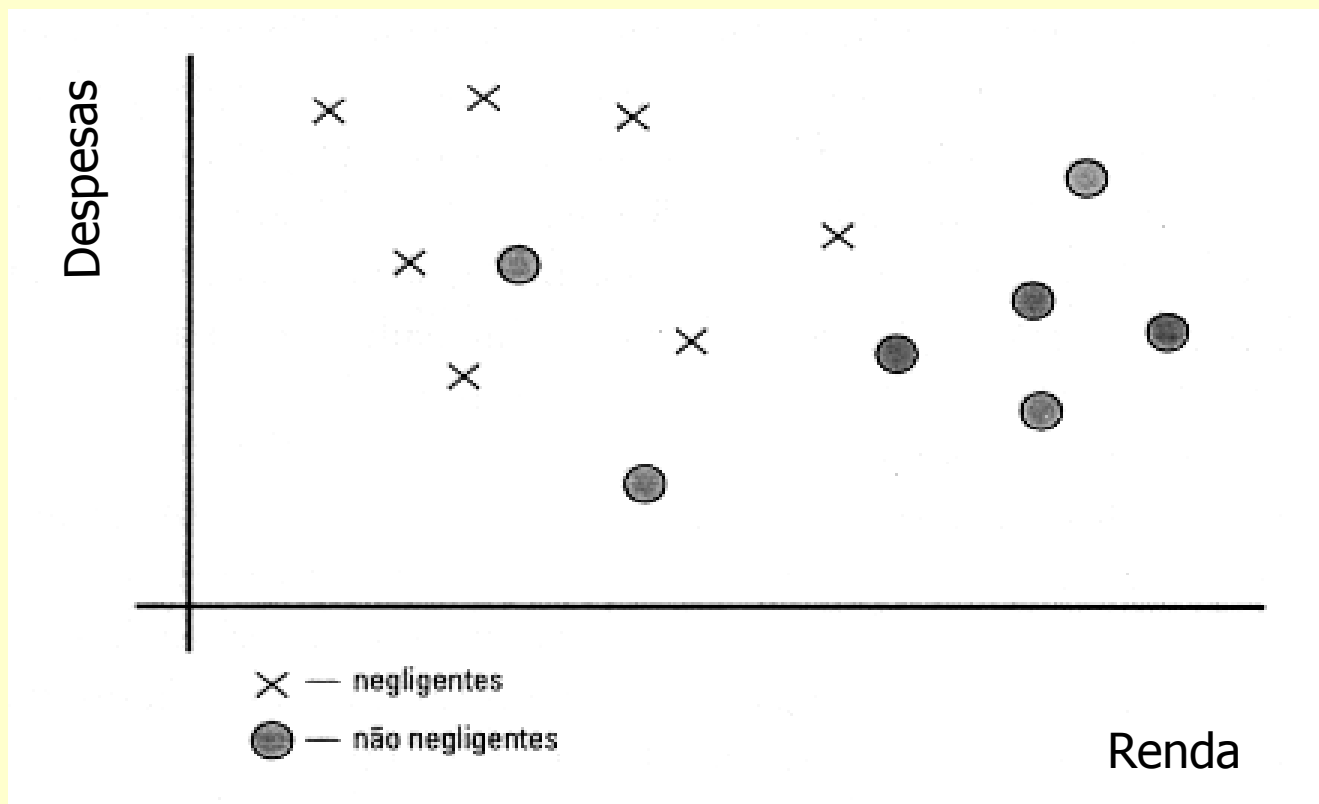
Então Crédito = SIM

Capacidade de Endividamento Mensal = 100*
(Renda Mensal - Despesas Mensais) / Renda Mensal

Renda Mensal, Despesas Mensais

1. INTRODUÇÃO E MOTIVAÇÃO

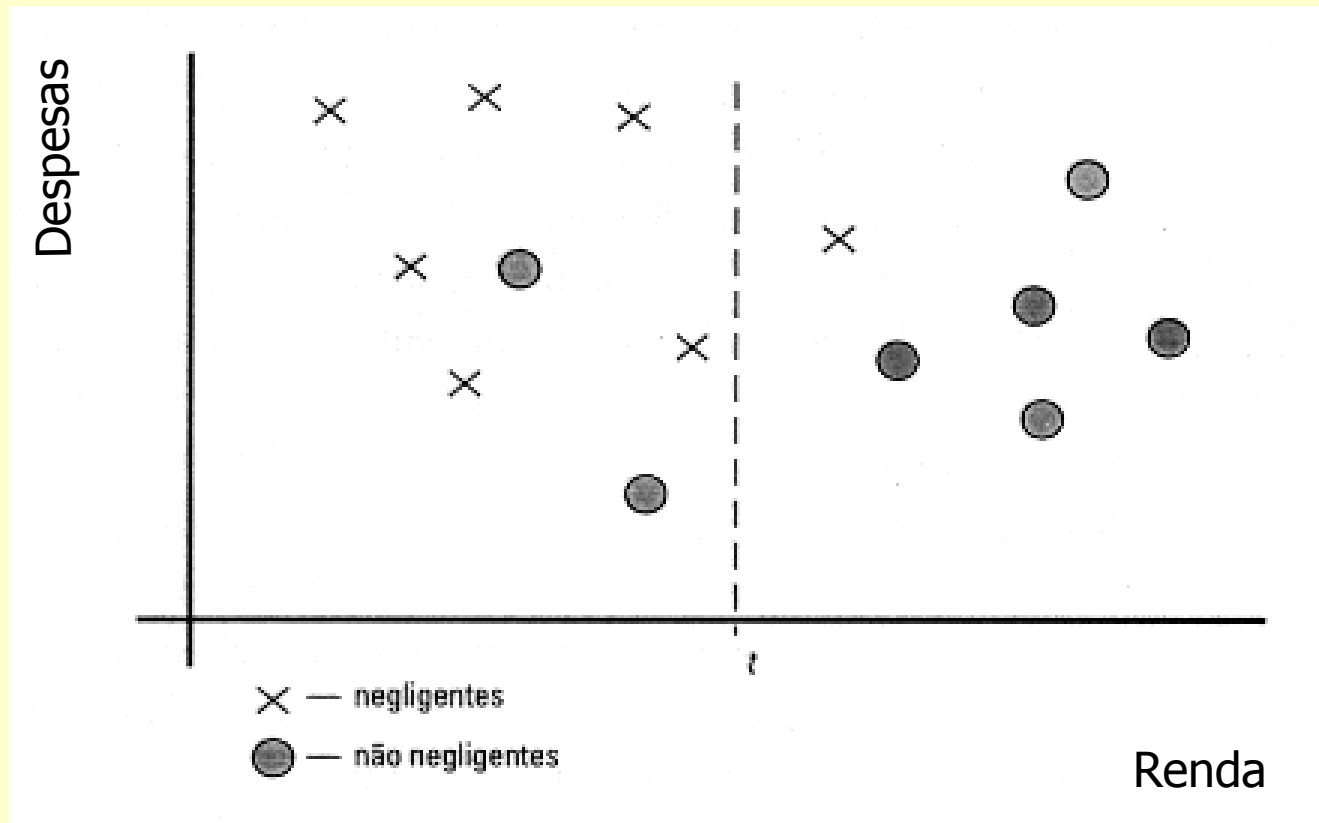
Exemplo de aplicação na área de concessão de crédito:
Conjunto de dados (Fatos)



1. INTRODUÇÃO E MOTIVAÇÃO

Exemplo de aplicação na área de concessão de crédito:

Padrão: Se renda > R\$ t Então Crédito = SIM (Cto)



1. INTRODUÇÃO E MOTIVAÇÃO

Outros Exemplos de Áreas de Aplicação:



Comércio



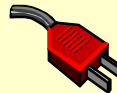
Finanças



Medicina



Educação



Energia



Telecomunicações



Meio-Ambiente



Indústria

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

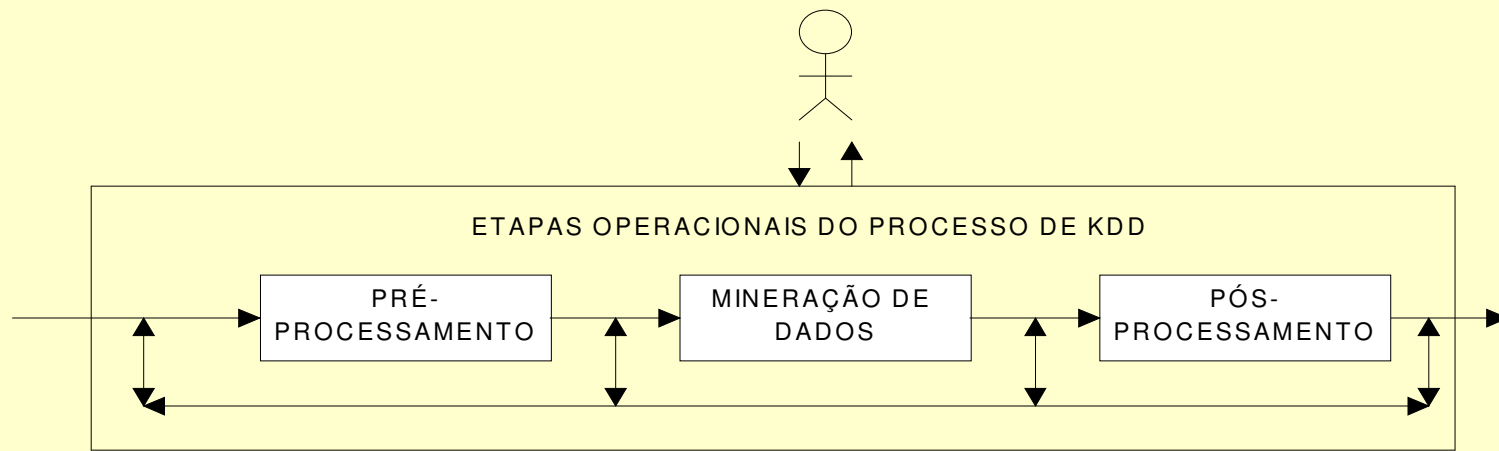
6. CONCLUSÕES

2. MINERAÇÃO DE DADOS E KDD

Processo de Descoberta do Conhecimento em Bases de Dados

- **KDD – Knowledge Discovery in Databases**

“É um **processo**, de várias etapas, não trivial, **interativo e iterativo**, para **identificação** de **padrões compreensíveis, válidos, novos** e potencialmente **úteis** a partir de bases de dados.” [Fayyad et al., 1996]



2. MINERAÇÃO DE DADOS E KDD

Exemplos de Operações de Pré-Processamento:

- Seleção dos dados
- Limpeza
- Codificação
- Enriquecimento

2. MINERAÇÃO DE DADOS E KDD

Seleção de Dados

- **Horizontal: escolha de casos**
 - **Amostragem**
 - **Segmentação do BD**
- **Vertical: escolha de características**
 - **Atributos relevantes**
 - **Redução de dimensionalidade**

2. MINERAÇÃO DE DADOS E KDD

Limpeza

- Verificação de consistência entre informações
- Correção de erros
- Preenchimento de valores desconhecidos
- Eliminação de informações redundantes
- Eliminação de valores não pertencentes ao domínio

Exemplo: Data de Nascimento

- Corretas nas seguradoras de vida;
- 30% a 40% em branco ou incorretas nos bancos;

2. MINERAÇÃO DE DADOS E KDD

Codificação

- Divide valores de atributos contínuos em intervalos codificados.

Ex: Renda

[0, 1000]	→ Faixa 1
[1001, 3000]	→ Faixa 2
[3001, 5000]	→ Faixa 3
etc...	

- Representa valores de atributos categóricos por códigos.

Ex: Sexo

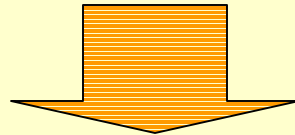
M	→ 1	F	→ 0
---	-----	---	-----

2. MINERAÇÃO DE DADOS E KDD

Enriquecimento - Exemplo: Perfil do Cliente

Atributos:

- Renda
- Despesas
- Tipo de Residência
- Bairro de Residência



Atributos:

- Renda
- Despesas
- Tipo de Residência
- Bairro de Residência
- Valor Médio Imóvel

2. MINERAÇÃO DE DADOS E KDD

Exemplos de Operações de Mineração de Dados:

- **CLASSIFICAÇÃO**
- **ASSOCIAÇÃO**
- **DETECÇÃO DE DESVIOS**
- **SUMARIZAÇÃO**
- **TEXT MINING**
- **Dentre muitas outras ...**

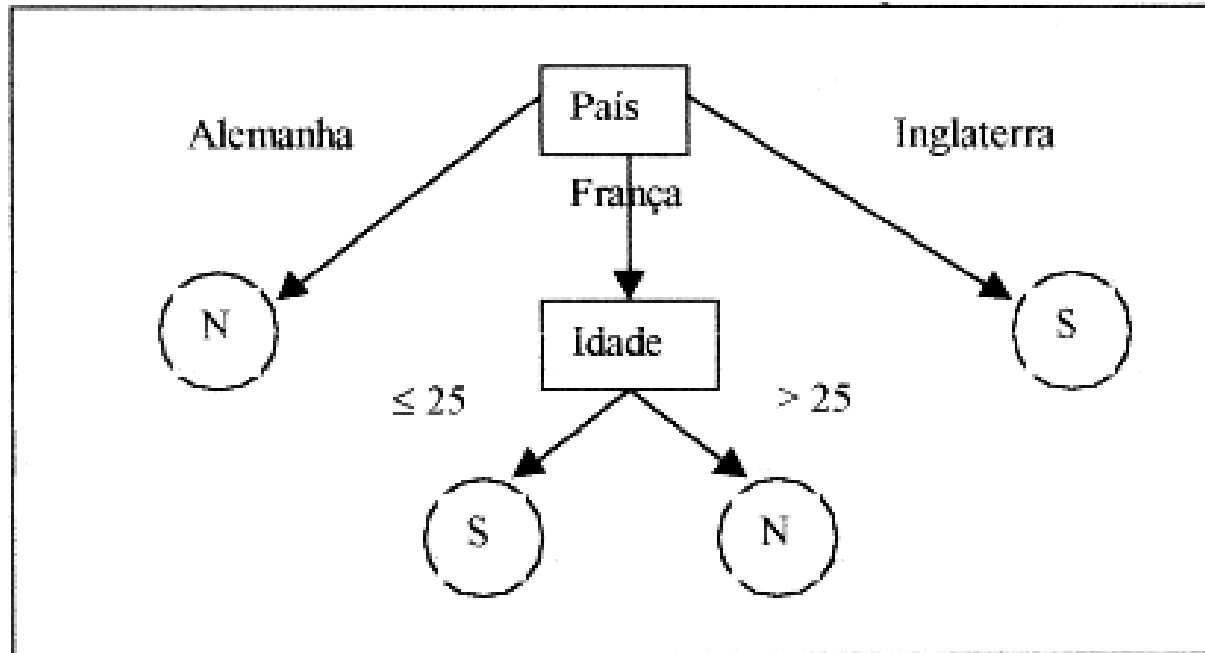
2. MINERAÇÃO DE DADOS E KDD

Exemplos de Operações de Pós-Processamento:

- **Construção de Árvores de Decisão**
- **Elaboração de Gráficos**
- **Elaboração de Relatórios Executivos**

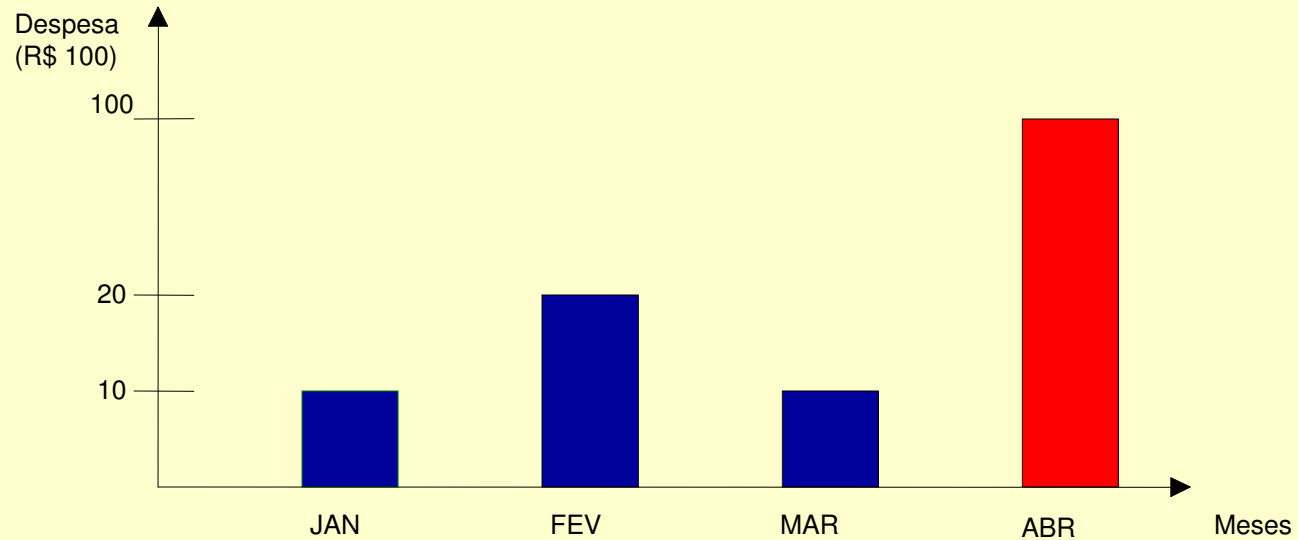
2. MINERAÇÃO DE DADOS E KDD

Construção de Árvores de Decisão



2. MINERAÇÃO DE DADOS E KDD

Elaboração de Gráficos



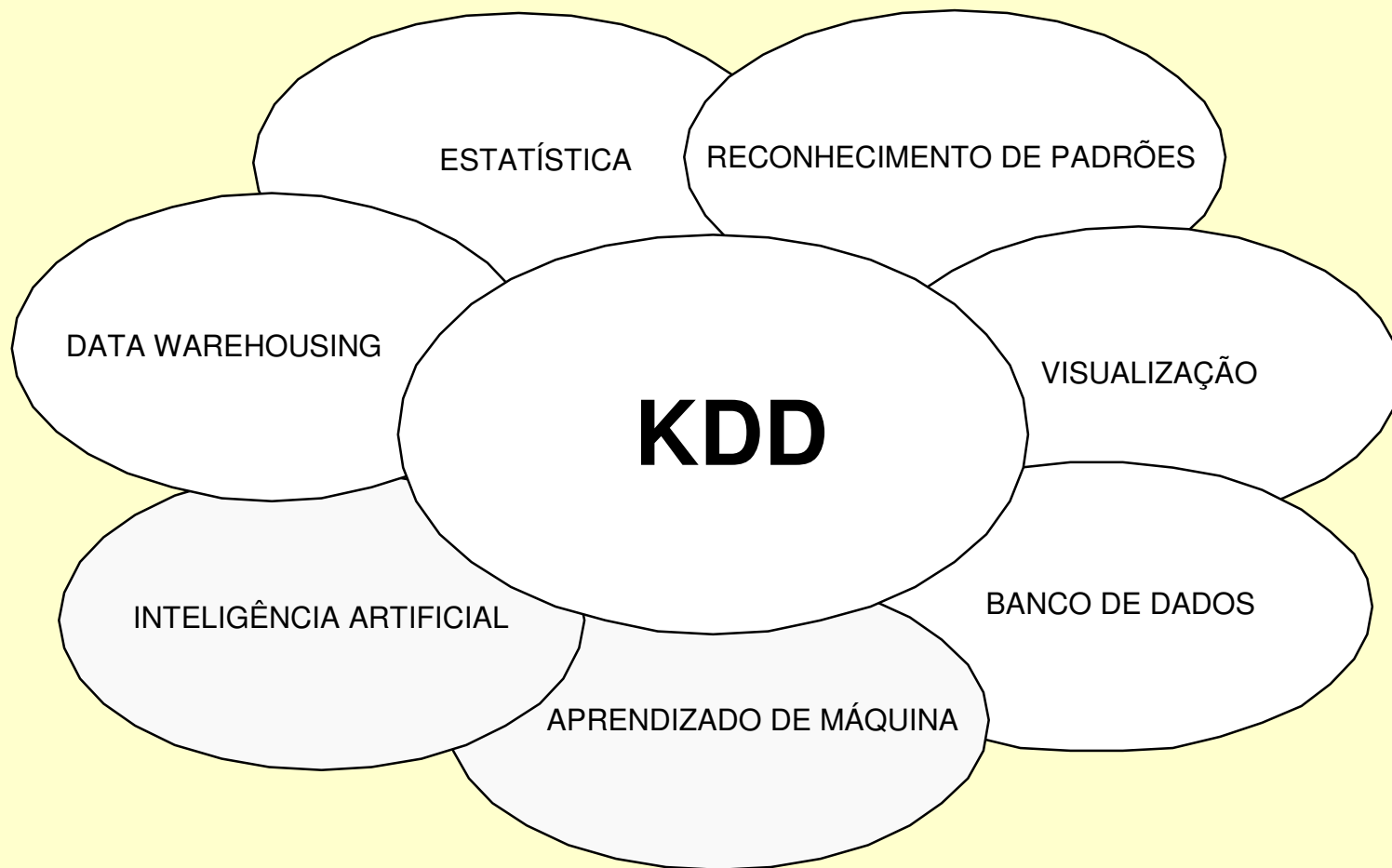
2. MINERAÇÃO DE DADOS E KDD

Elaboração de Relatórios Executivos

- Se (País = Alemanha) Então Comprar = Não
- Se (País = Inglaterra) Então Comprar = Sim
- Se (País = França e Idade ≤ 25) Então Comprar = Sim
- Se (País = França e Idade > 25) Então Comprar = Não

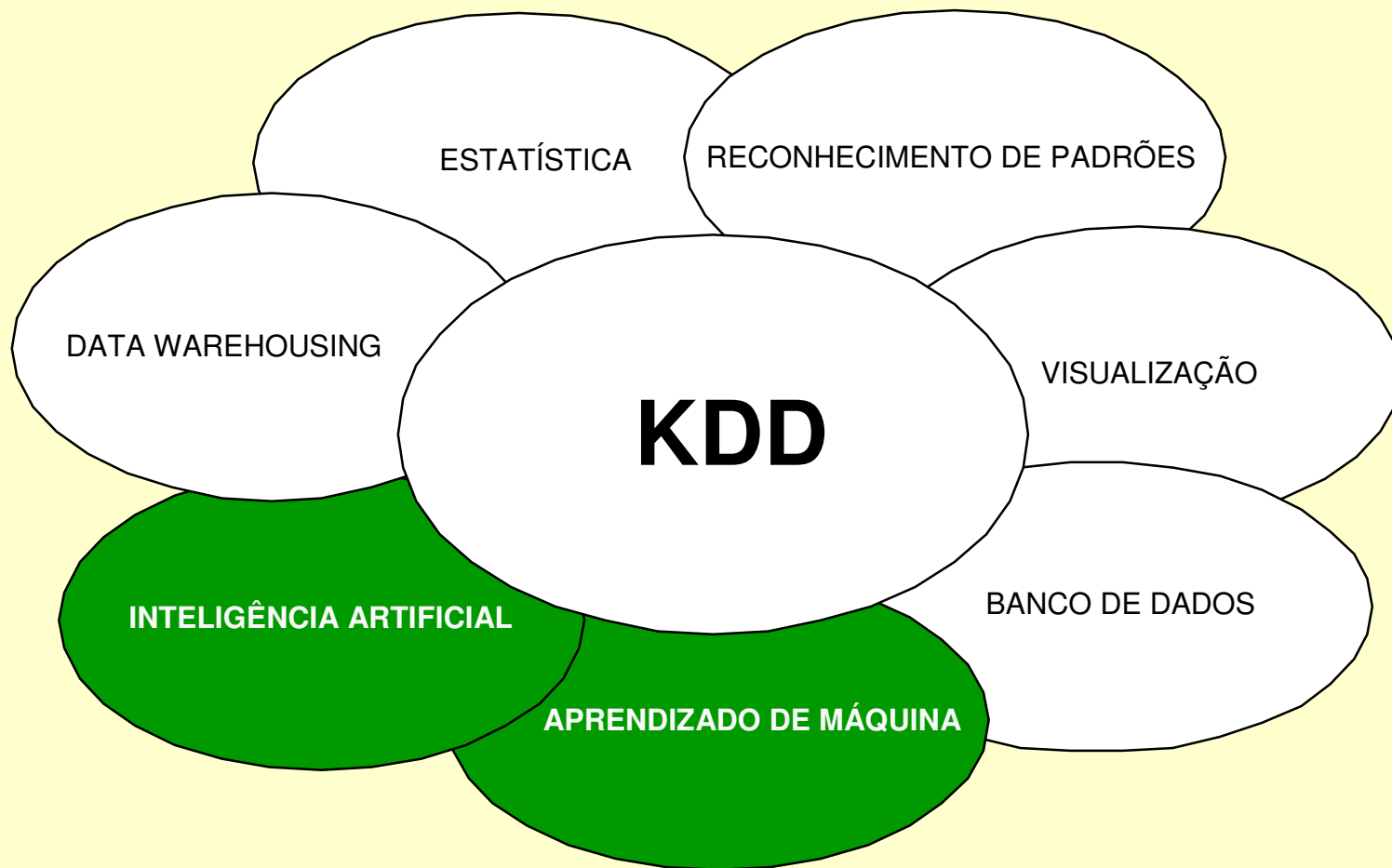
2. MINERAÇÃO DE DADOS E KDD

Áreas de Origem:



2. MINERAÇÃO DE DADOS E KDD

Áreas de Origem:



2. MINERAÇÃO DE DADOS E KDD

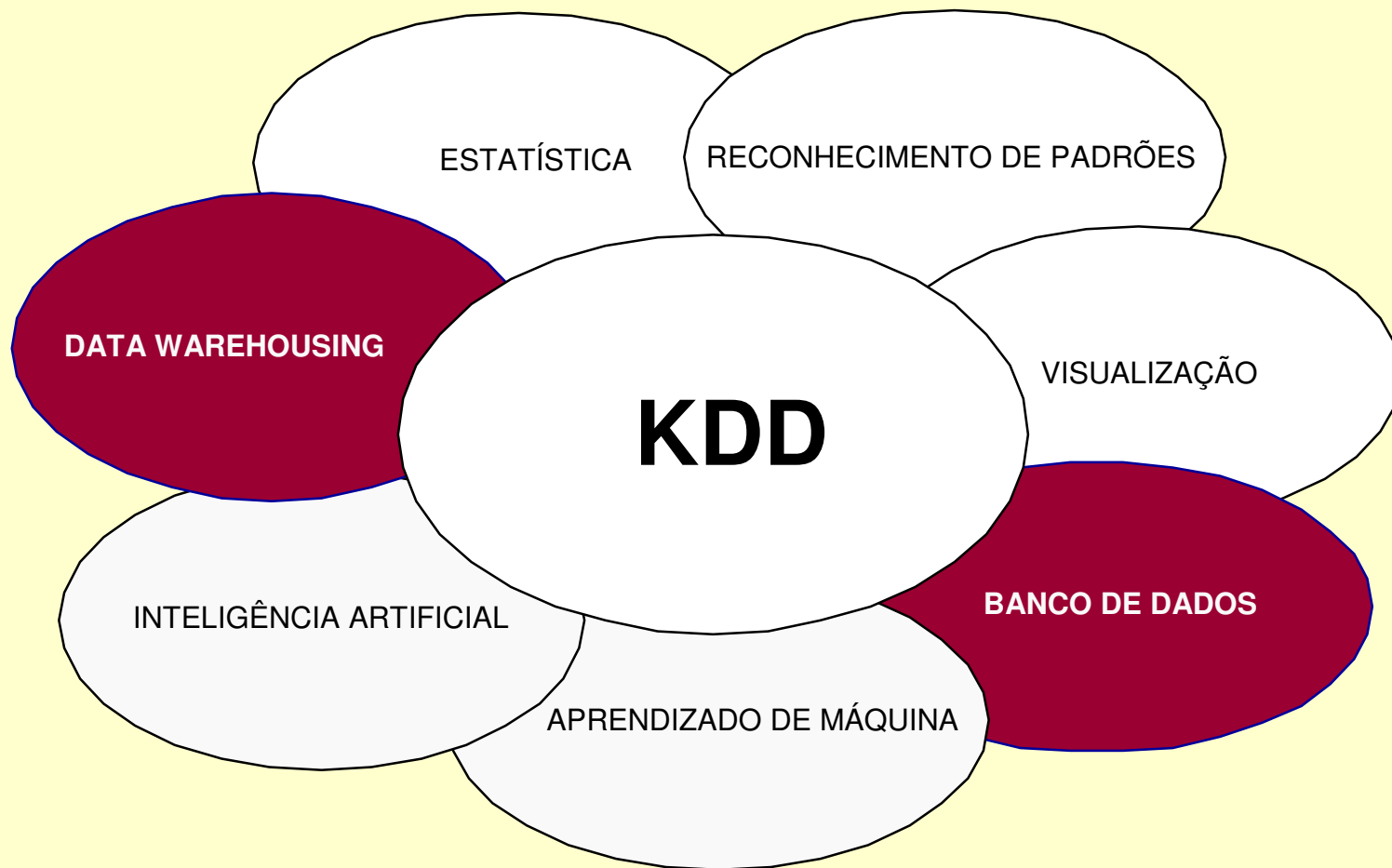
Áreas de Origem:

Aprendizado de Máquina - Inteligência Artificial:

- **Redes Neurais**
- **Algoritmos Genéticos**
- **Lógica Nebulosa**
- **Lógica Indutiva**
- **Árvores de Decisão**

2. MINERAÇÃO DE DADOS E KDD

Áreas de Origem:



2. MINERAÇÃO DE DADOS E KDD

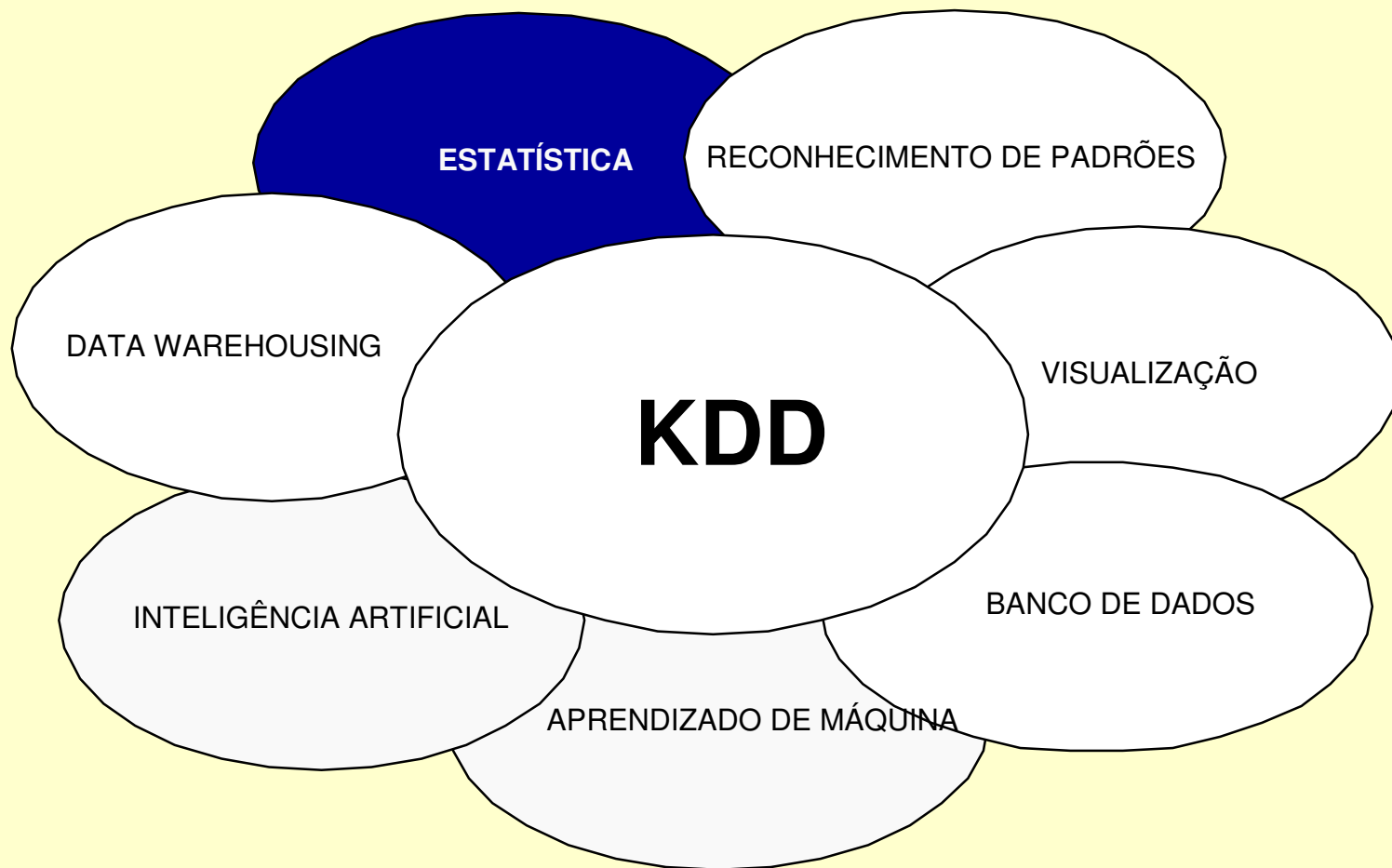
Áreas de Origem:

Banco de Dados / Data Warehouses:

- **Data Warehousing**
- **SQL**
- **OLAP**
- **DMQL**

2. MINERAÇÃO DE DADOS E KDD

Áreas de Origem:



2. MINERAÇÃO DE DADOS E KDD

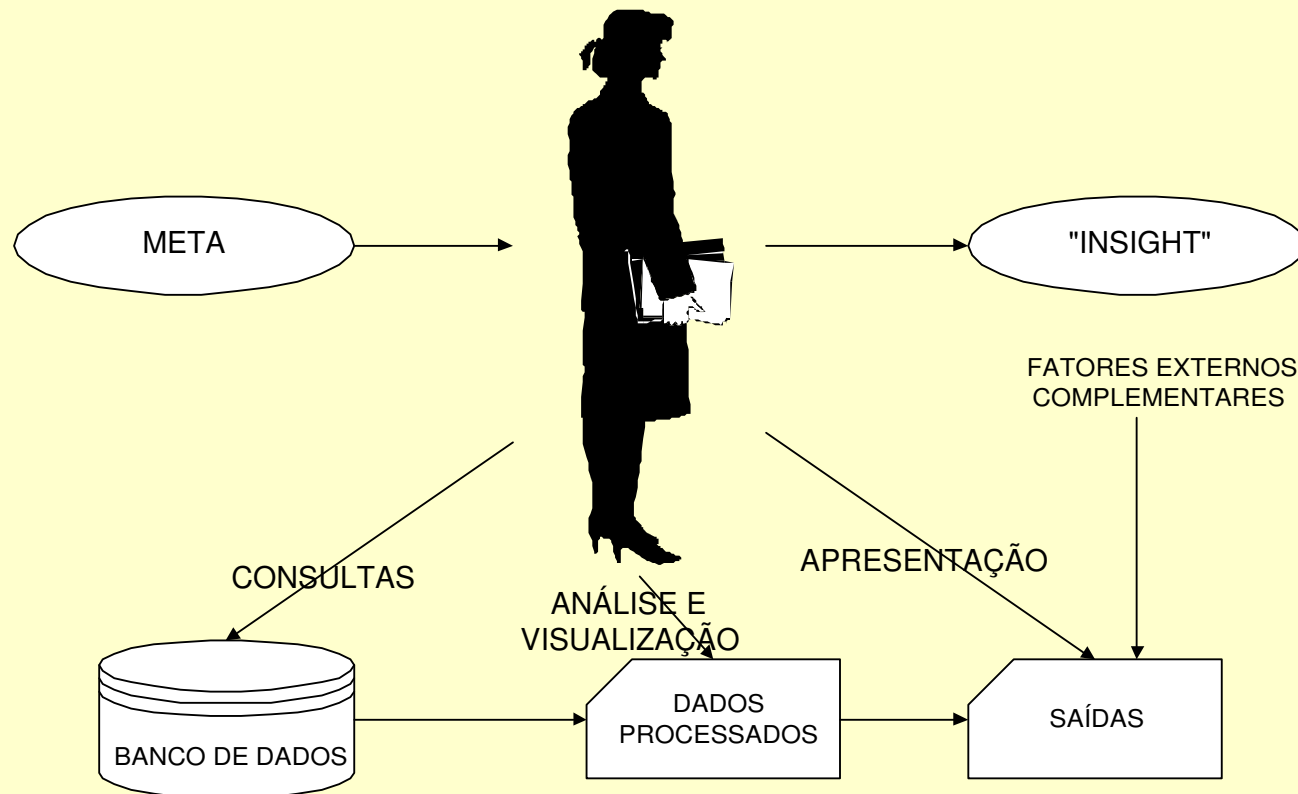
Áreas de Origem:

Estatística:

- **Classificadores Bayesianos**
- **Redes Bayesianos**
- **EDA - Exploratory Data Analysis**

2. MINERAÇÃO DE DADOS E KDD

A importância do usuário no processo KDD



2. MINERAÇÃO DE DADOS E KDD

Gerações da Mineração de Dados [Piatetsky-Shapiro, 2001]

- 1ª Geração
 - Anos 90
 - Ferramentas de Pesquisa voltadas a uma única tarefa, sem suporte às demais etapas de KDD
 - Exemplos: c4.5, Rede Neural, Autoclass, etc...

2. MINERAÇÃO DE DADOS E KDD

Gerações da Mineração de Dados [Piatetsky-Shapiro, 2001]

- 2ª Geração
 - Meados dos anos 90
 - Ferramentas chamadas “suites”: Pacote para aplicação com suporte ao pré-processamento e à visualização
 - Requerem conhecimento significativo da teoria estatística
 - Exemplos SPSS, Intelligent Miner, SAS, etc...

2. MINERAÇÃO DE DADOS E KDD

Gerações da Mineração de Dados [Piatetsky-Shapiro, 2001]

- **3ª Geração**
 - **Final dos anos 90**
 - **Soluções orientadas à resolução de problemas específicos em empresas**
 - **Possuem interfaces orientadas aos usuários**
 - **Escondem a complexidade da MD**
 - **Exemplos: Falcon (Detecção Fraude em Cartão)**

2. MINERAÇÃO DE DADOS E KDD

Macro-Objetivos da Mineração de Dados [Zaki, 2002]:

- **Predição: Histórico x Novas Situações**
- **Descrição: Modelo Descritivo do Conhecimento**

Orientação das Tarefas de Mineração de Dados [Zaki, 2002]:

- **Para Verificação: Hipótese Postulada x Validação**
- **Para Descoberta: Extração de novos conhecimentos**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

- **CLASSIFICAÇÃO**
- **ASSOCIAÇÃO**
- **DETECÇÃO DE DESVIOS**
- **SUMARIZAÇÃO**
- **TEXT MINING**

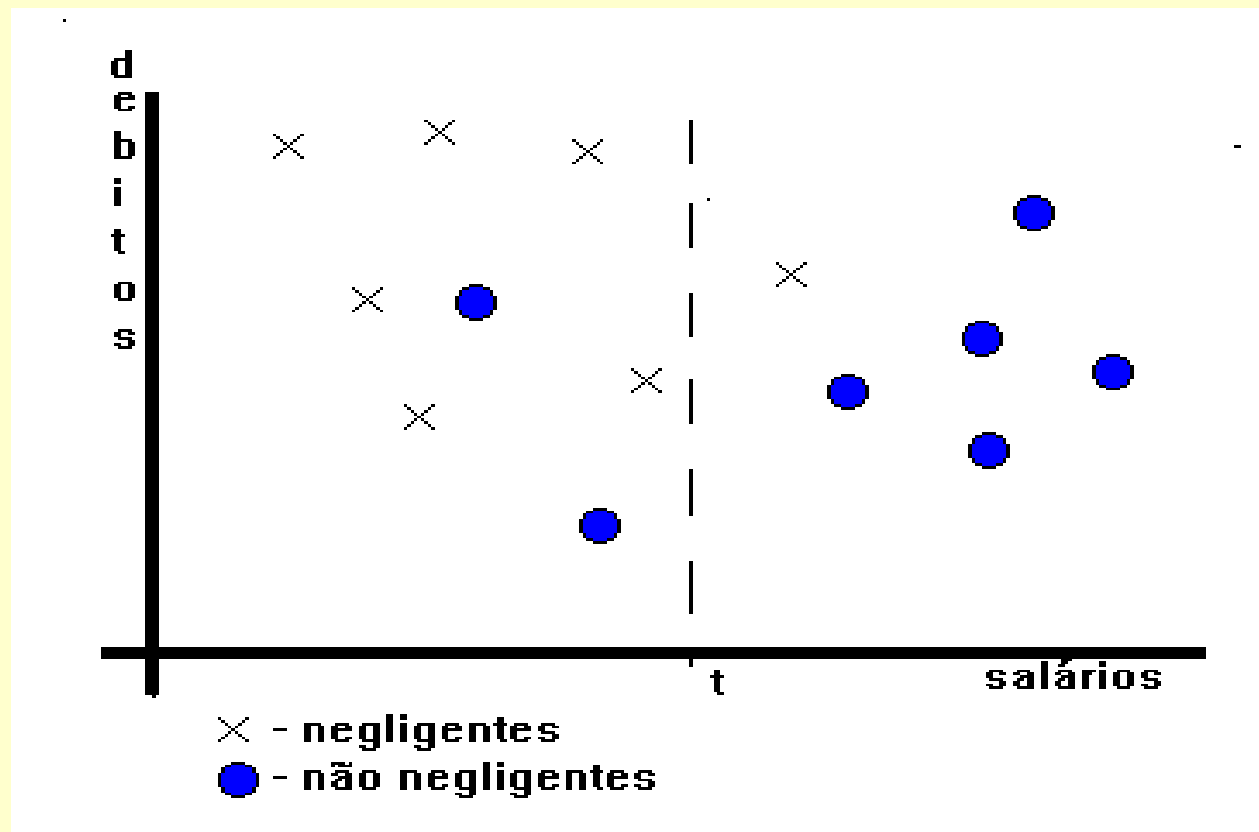
3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

- **CLASSIFICAÇÃO**
- **ASSOCIAÇÃO**
- **DETECÇÃO DE DESVIOS**
- **SUMARIZAÇÃO**
- **TEXT MINING**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: CLASSIFICAÇÃO

EXEMPLO



3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: CLASSIFICAÇÃO

EXEMPLOS DE TÉCNICAS TRADICIONAIS:

- **REDES NEURAIS → BACKPROPAGATION**
- **ÁRVORES DE DECISÃO → ID3, C4.5**
- **ALGORITMOS GENÉTICOS → RULE EVOLVER**
- **ESTATÍSTICA → CLASSIFICADORES BAYESIANOS**
- **BASEADAS EM INSTÂNCIA → K-NN**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: CLASSIFICAÇÃO

EXEMPLOS DE APLICAÇÕES

- **FINANÇAS E INVESTIMENTOS**
 - **SEGUROS**
- **RECONHECIMENTO DE IMAGEM**
- **RECONHECIMENTO DE VOZ**
 - **ETC...**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

- CLASSIFICAÇÃO
- ASSOCIAÇÃO
- DETECÇÃO DE DESVIOS
- SUMARIZAÇÃO
- TEXT MINING

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: ASSOCIAÇÃO

“Consiste em encontrar **conjuntos de itens** que **ocorram simultaneamente de forma frequente** em um banco de dados.”

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: ASSOCIAÇÃO

EXEMPLO

Encontrar produtos frequentemente vendidos de forma conjunta

N. Trans.	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
1	não	sim	não	sim	sim	não	não
2	sim	não	sim	sim	sim	não	não
3	não	sim	não	sim	sim	não	não
4	sim	sim	não	sim	sim	não	não
5	não	não	sim	não	não	não	não
6	não	não	não	não	sim	não	não
7	não	não	não	sim	não	não	não
8	não	não	não	não	não	não	sim
9	não	não	não	não	não	sim	sim
10	não	não	não	não	não	sim	não

Algumas Regras de Associação:

Café $\rightarrow$ Pão

Café $\wedge$ Pão $\rightarrow$ Manteiga

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: ASSOCIAÇÃO

EXEMPLOS DE ALGORITMOS TRADICIONAIS:

- **APRIORI**
- **DHP – DIRECT HASHING AND PRUNING**
- **PARTITION**
- **DIC – DYNAMIC ITEMSET COUNTING**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: ASSOCIAÇÃO

EXEMPLOS DE APLICAÇÕES

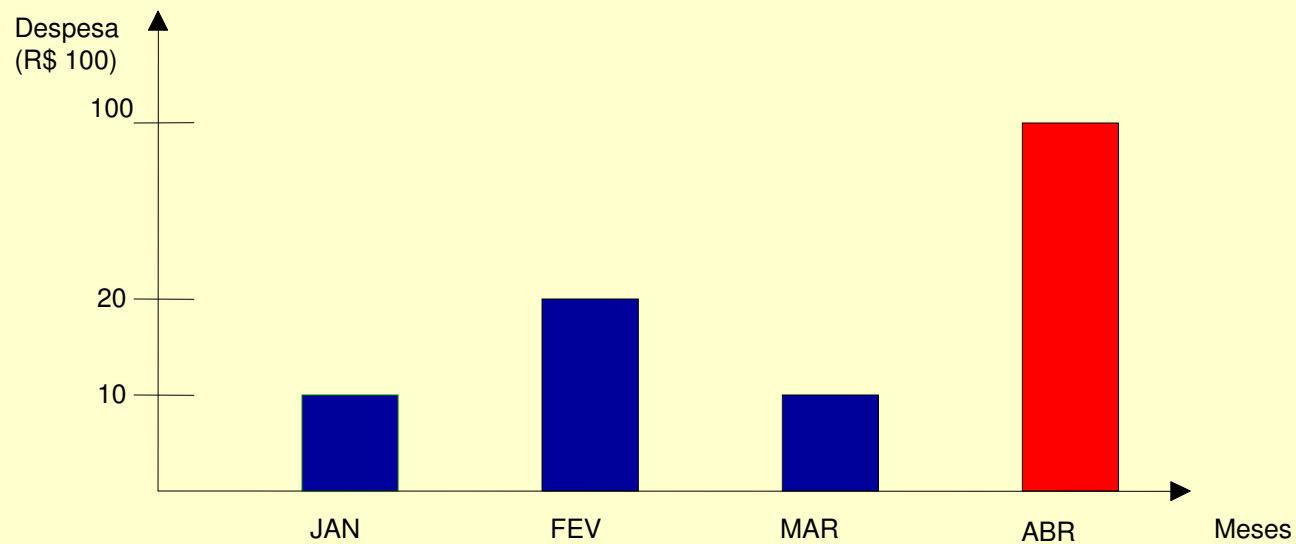
- **MARKETING**
- **PESQUISAS CIENTÍFICAS – PADRÕES SIMULTÂNEOS**
- **CLASSIFICAÇÃO POR REGRAS DE ASSOCIAÇÃO**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

- CLASSIFICAÇÃO
- ASSOCIAÇÃO
- DETECÇÃO DE DESVIOS
- SUMARIZAÇÃO
- TEXT MINING

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

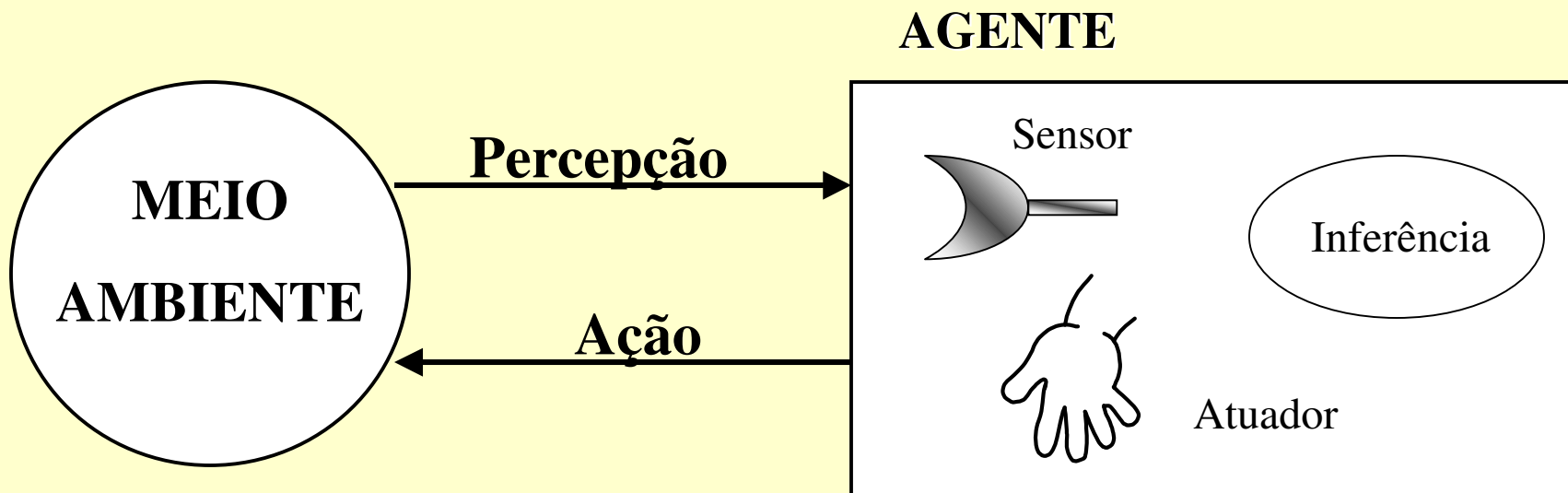
TAREFA: DETECÇÃO DE DESVIOS



3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: DETECÇÃO DE DESVIOS

TÉCNICA: AGENTES INTELIGENTES [Chan, 2001]



3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: DETECÇÃO DE DESVIOS

EXEMPLOS DE APLICAÇÕES

- **DETECÇÃO DE FRAUDES:**

CARTÃO DE CRÉDITO

PLANOS DE SAÚDE

ARRECADAÇÃO

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

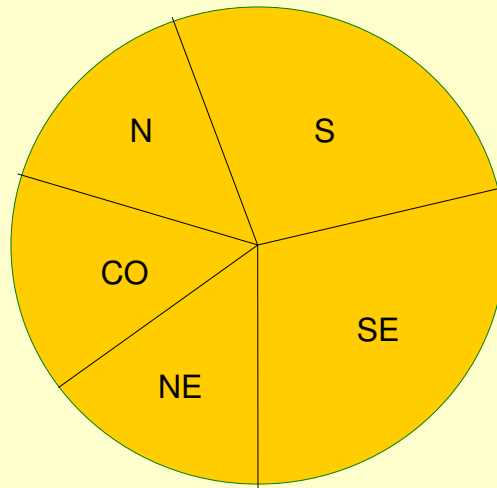
- **CLASSIFICAÇÃO**
- **ASSOCIAÇÃO**
- **DETECÇÃO DE DESVIOS**
- **SUMARIZAÇÃO**
- **TEXT MINING**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: SUMARIZAÇÃO

“Consiste em **descrever** as **características** de **subconjuntos** da base de **dados**.”

Ex: Distribuição dos Assinantes da Revista “X” por Regiões.



3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: SUMARIZAÇÃO

EXEMPLOS DE ALGORITMOS TRADICIONAIS:

- **MODELOS ESTATÍSTICOS – VISUALIZAÇÃO**
 - **CUBOS DE DADOS - VISUALIZAÇÃO**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

- **CLASSIFICAÇÃO**
- **ASSOCIAÇÃO**
- **DETECÇÃO DE DESVIOS**
- **SUMARIZAÇÃO**
- **TEXT MINING**

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: TEXT MINING

- Descobrir conhecimento em textos em Linguagem Natural
- Utiliza recursos de PLN - Processamento de Linguagem Natural

Exemplos de Aplicações:

- Recuperação de Informação
- Extração de Dados
- Classificação de Documentos
- Resumo / Correção de Textos

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

TAREFA: TEXT MINING

TÉCNICA: ASSOCIAÇÃO DE TERMOS

Ex.: “Muitas aplicações de KDD utilizam Data Warehouses”.

Após PLN: Aplicações, KDD, Data Warehouses

Após MD: KDD e Data Warehouses são termos frequentes:
têm alguma relação entre si.

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

6. CONCLUSÕES

4. EJEMPLOS DE HERRAMIENTAS

- SAS – Enterprise Miner
- SPSS
- PolyAnalist
- Clementine
- Intelligent Miner
- **WizRule e WizWhy**
- **Bramining**
- **Rule Evolver**

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

6. CONCLUSÕES

5. EXEMPLOS DE APLICAÇÕES

Áreas de Aplicação:



Comércio



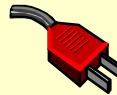
Finanças



Medicina



Educação



Energia



Telecomunicações



Meio-Ambiente



Indústria

5. EXEMPLOS DE APLICAÇÕES

- **Comércio / Marketing**

Perfil do Consumidor (Marketing Direto), Promoção de Produtos, Segmentação de Mercado, etc;...

- **Finanças**

Análise de Investimentos, Análise de Crédito, Detecção de Fraudes em compras de Cartão de Crédito, etc;...

- **Medicina**

Diagnóstico e Prevenção de Doenças, Detecção de Fraudes em Planos de Saúde, etc;...

5. EXEMPLOS DE APLICAÇÕES

- **Educação**

Análise de Matrículas e Demandas por Escolas,
Evasão Escolar;...

- **Energia**

Previsão de Demanda, Distribuição de Recursos;...

- **Telecomunicações**

Detecção de falhas, Dimensionamento de Sistemas de
Comunicação, Detecção de Fraudes;...

5. EXEMPLOS DE APLICAÇÕES

- **Meio Ambiente**

Monitoramento ambiental, Prevenção de desequilíbrios ecológicos;...

- **Indústria**

Previsão de demanda, Planejamento da produção e distribuição;...

- **Área Social**

Caracterização de Perfil para Reintegração Social;...

ROTEIRO

1. INTRODUÇÃO E MOTIVAÇÃO

2. MINERAÇÃO DE DADOS E KDD

3. TAREFAS, ALGORITMOS E TÉCNICAS DE KDD

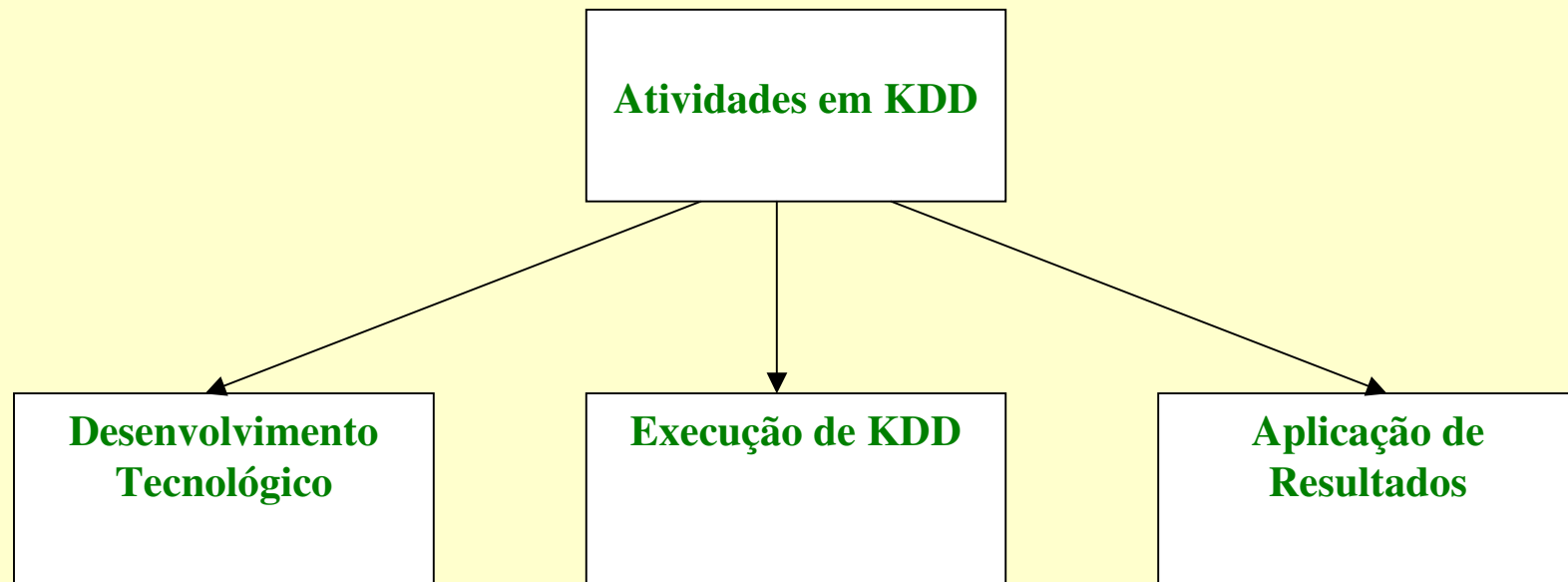
4. EXEMPLOS DE FERRAMENTAS

5. EXEMPLOS DE APLICAÇÕES

6. CONCLUSÕES

6. CONCLUSÕES

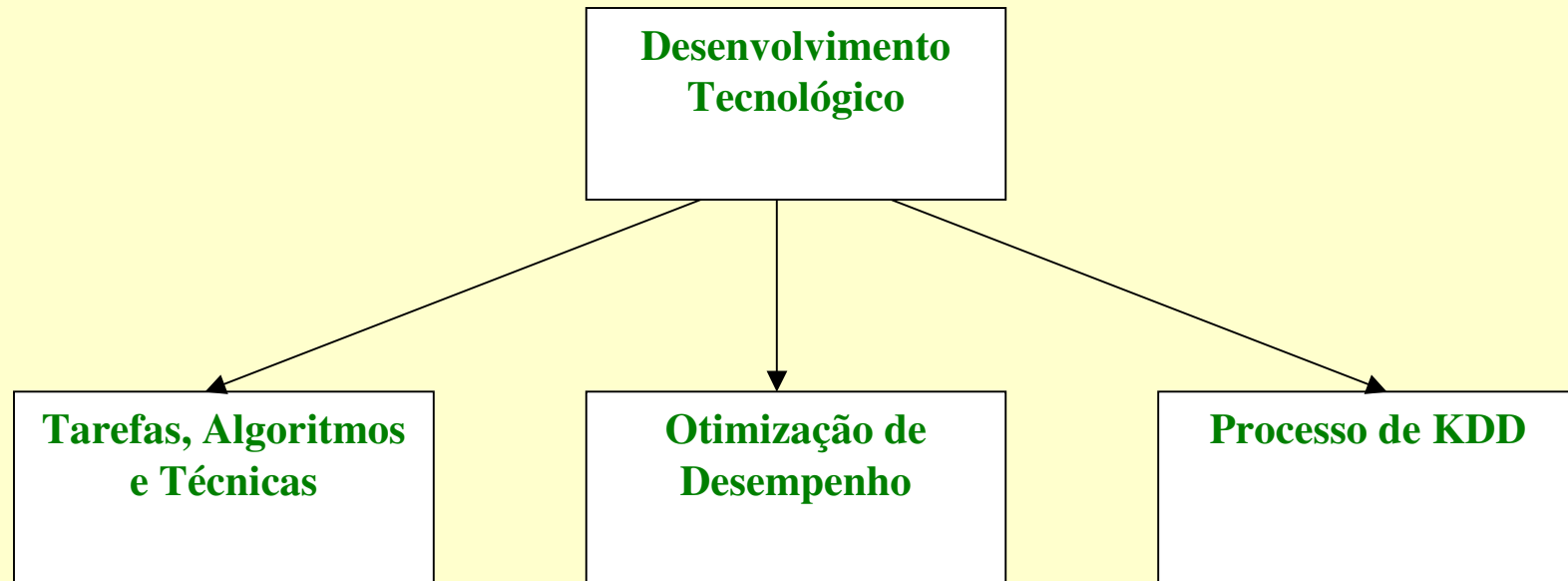
Atividades em KDD - uma Taxonomia



[Goldschmidt et al., 2002a]

6. CONCLUSÕES

Atividades em KDD - uma Taxonomia



[Goldschmidt et al., 2002a]

6. CONCLUSÕES

Tendências e Perspectivas:

- **Mineração de Dados em Larga Escala**
- **Mineração de Dados Paralela e Distribuída**
- **Mineração de Dados Ativa**
- **Mineração de Dados na WEB**
- **Assistência ao Controle do Processo de KDD**