



# Tecnologias e Sistemas de Informação

---

## Sistemas Inteligentes

### Aula 1 - Introdução a Sistemas Inteligentes e a Mineração de Dados

Prof. José Artur Quilici-Gonzalez  
Email: [jose.gonzalez@ufabc.edu.br](mailto:jose.gonzalez@ufabc.edu.br)



# Roteiro

---

**Introdução**



**Dado, Informação e Conhecimento**



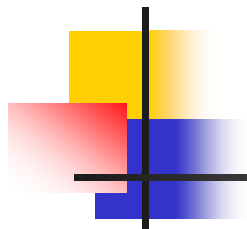
**Descoberta de Conhecimento em Bases de Dados**



**Mineração de Dados**



**Referência Bibliográfica**



# INTRODUÇÃO



# Introdução

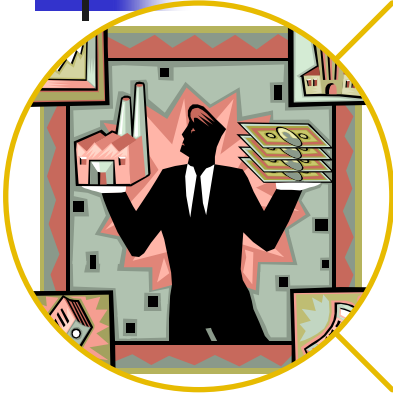
---

O que se entende por **Sistemas Inteligentes ??**

O que é um **Sistema ???**

E o que é **Inteligente ????**

# Sistema



**Sistema:** do grego, significa "combinar", "ajustar", "formar um conjunto articulado"



Um **sistema** pode ser formado por um conjunto de pessoas, recursos, instalações e métodos direcionados a **atingir um fim**

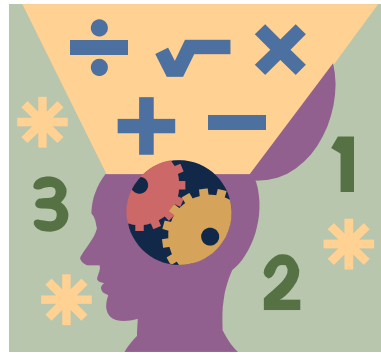
O **sistema financeiro** compreende pessoas, regras, prédios etc. combinados para realizar operações financeiras

O **sistema respiratório** reúne vários órgãos responsáveis pelo fornecimento de oxigênio para as células

# Inteligente

**Inteligência:** é muito difícil definir!

Seria a **síntese de conhecimentos** para atingir um objetivo identificável?



É melhor tentar entender **Comportamento Inteligente** como a habilidade para resolver novos problemas ou capacidade de adaptar-se para melhorar seu desempenho

# Sistema Inteligente

Um  
**Sistema  
Inteligente**  
pode ser



um **sistema físico** - essencialmente hardware com capacidade adaptativa

um **sistema computacional** - predominantemente um software que comanda uma máquina

um **sistema híbrido** - um robô com sensores, atuadores, sistema operacional, programas computacionais etc.

# Sistema Inteligente



Independentemente da natureza do **Sistema Inteligente**, ele terá que demonstrar habilidade para **adaptar-se a mudanças** em seu ambiente não previstas pelo projetista

e tomar decisões baseadas em novos conhecimentos adquiridos com experiência própria





---

# Dado, Informação e Conhecimento



# Era da Informação

---

Pode parecer paradoxal afirmar que em plena “Era da Informação” não haja uma visão clara do que é informação

Convém, no entanto, lembrar que este mesmo cenário já ocorreu em outros períodos da história humana

A genética mendeliana do século XIX, por exemplo, descobriu experimentalmente que certas características das ervilhas poderiam ser transmitidas através da hibridação das plantas

Mas foi preciso esperar por um século até a descoberta da estrutura do DNA para que uma teoria robusta explicasse o mecanismo da herança genética

Antes disso, as explicações dos fenômenos de transmissão de herança genética eram meramente funcionais



# Dado

---

- É importante definir **Dado**, **Informação** e **Conhecimento** porque eles estão na base do **Aprendizado**
- **Dado** é um fato registrado ou uma quantidade (ou qualidade) conhecida sem a necessidade de elaboração
- Por exemplo:
  - O peso (ou massa!) de uma pessoa
  - O valor da dívida de um cliente
  - Os registros de temperatura dos últimos anos
  - Os conceitos ou as notas de provas de alunos
  - O nome de um empregado etc.





# Informação



<http://www>

- Quando os **dados** apresentam alguma relação entre si, eles podem ser **contextualizados**, adquirindo um **significado**
- **Informação envolve dados contextualizados**
- Associando-se o peso (ou a massa) de uma pessoa à sua altura, obtém-se o Índice de Massa Corporal (IMC)
  - Nesse contexto, peso e altura são dados, IMC é informação
- Apenas o valor da dívida de um cliente não permite a um gerente de banco decidir sobre um pedido de empréstimo
  - É necessário levar em conta a renda do cliente, ou talvez o histórico de suas movimentações, para que uma interpretação confiável possa ser feita



# Conhecimento

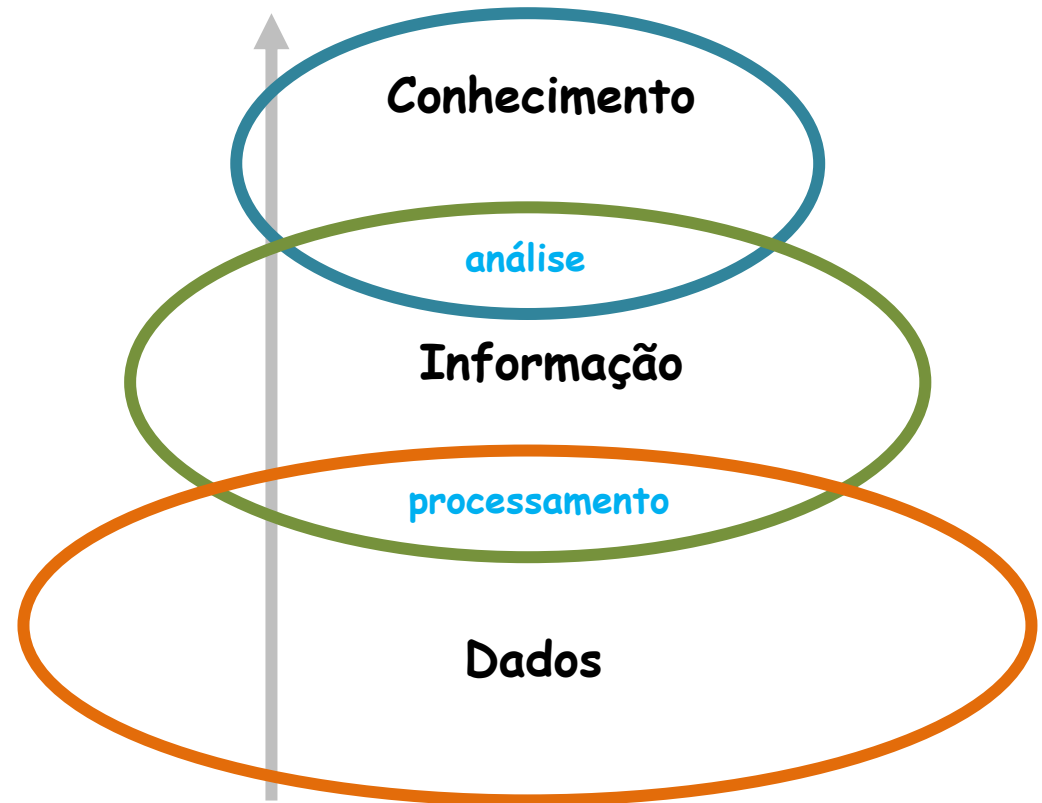
---



- As informações úteis a um propósito permitem ao indivíduo compreender uma situação, ou formar uma crença justificada sobre um fato
- Portanto, o conhecimento se forma a partir das informações necessárias para o **entendimento** de uma situação
- Nesse contexto, **conhecimento** é o resultado da análise das informações relacionadas a um fato ou evento, ou ainda
- **Conhecimento** é a percepção de como certa informação pode ajudar na realização de uma tarefa específica

# Dados, Informação e Conhecimento

Embora a diferenciação entre **Dados**, **Informação** e **Conhecimento** seja tênue, é comumente aceito que estes três conceitos podem ser relacionados em uma estrutura hierárquica





# Desempenho do Sistema

---

- O **desempenho** de um sistema pode ser aferido comparando-se o **resultado obtido com o esperado**, ou com os **recursos empregados para chegar àquele resultado**
- Quando alguém toma um remédio para tratar-se de uma doença, e se cura, geralmente dizemos que tal remédio foi **eficaz** porque o resultado obtido se aproximou do esperado
  - Neste caso, para medir o desempenho do sistema, comparamos sua saída com o resultado esperado
- Por outro lado, quando dizemos que a **eficiência** de um motor a combustão está em torno de 30%, estamos comparando a energia de entrada (combustível) com a de saída (potência mecânica)
  - Agora o desempenho foi caracterizado pela relação entre entrada e saída do sistema

# Gerenciamento do Conhecimento



Nas últimas décadas foram criados vários sistemas para otimizar cada etapa do processo de **extração de conhecimento**



Esses sistemas cuidam desde a coleta de dados operacionais de uma organização, passando pela **Mineração de Dados**, até a análise dos resultados



O processo para a extração de conhecimento é chamado de **Descoberta de Conhecimento em Bases de Dados**, ou **Knowledge Discovery in Databases**, ou **KDD**





---

# DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

# Modelo Ideal X Sistema Real



**Modelos ideais** podem não resolver plenamente problemas de **sistemas reais**, porque dados reais geralmente apresentam redundâncias, ausências de valores, erros, inconsistências etc.



**Processamento de dados** geralmente requer **pré-processamento**, pois a qualidade dos dados de entrada pode ter um impacto significativo sobre a etapa seguinte, a de **Mineração de Dados**

# Representação do Conhecimento



O que se espera com a **Mineração de Dados** é obter conhecimento, ou uma **representação de conhecimento** na forma de regras ou de estruturas equivalentes, que oriente uma decisão



E, com a representação do conhecimento em mãos, há a necessidade de um **pós-processamento** para interpretar e validar os resultados obtidos



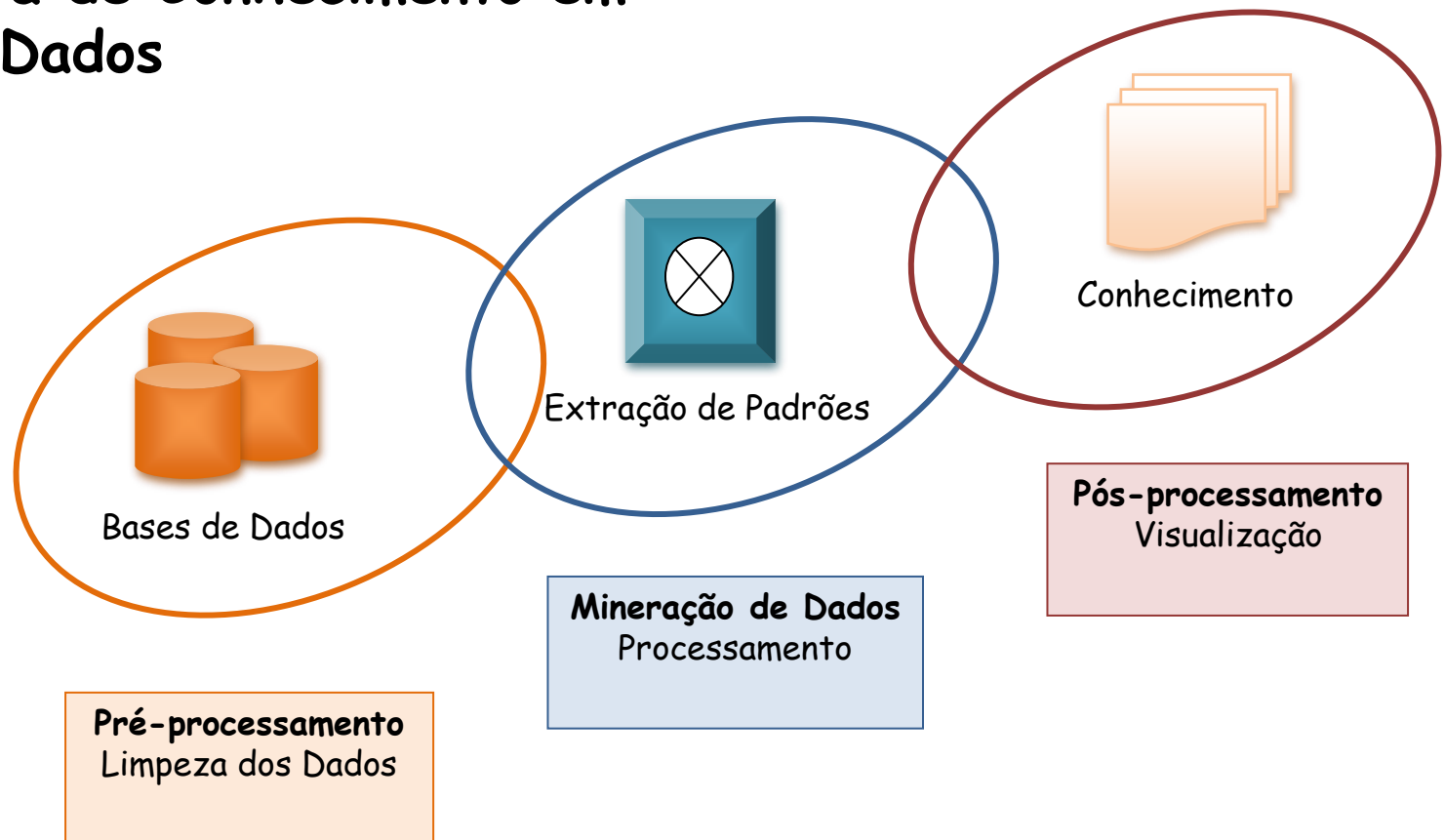
**Sistemas Especialistas** utilizados para diagnóstico médico precisam expor de forma inteligível a um especialista todas as etapas de inferência que levaram àquele resultado



Caso contrário, o médico poderá não se sentir seguro para acolher o diagnóstico produzido pelo Sistema Especialista

# Descoberta de Conhecimento

Principais etapas do processo de  
**Descoberta de Conhecimento em**  
**Bases de Dados**



# Pré-Processamento

O propósito do **Pré-processamento** é eliminar eventuais problemas nos dados brutos e colocá-los num formato apropriado para a etapa seguinte, a da **Mineração de Dados**

O Pré-processamento pode **melhorar** significativamente a **eficiência dos algoritmos** que serão usados

Os **dados originais** podem ter sido coletados e reunidos por diferentes departamentos, apresentando **valores espúrios ou ausentes**, ou contendo **redundâncias**

**Principais Transformações:**  
Limpeza e Fusão dos Dados Brutos, Eliminação de Ruídos e Redundâncias, Redução do Número de Variáveis



# Mineração de Dados

---

O objetivo é descobrir de forma automatizada relações ou padrões implícitos em grandes quantidades de dados

Ou comprovar alguma hipótese a partir de informações até então não facilmente perceptíveis nos dados



# Pós-Processamento

E se os resultados obtidos não forem os esperados?

Na etapa de *Pós-processamento* é possível visualizar e interpretar as regras ou os padrões obtidos com a Mineração de Dados, e eliminar os resultados equivocados ou pouco representativos

Nesta fase é bem comum a aplicação de testes estatísticos para validação dos resultados

Dependendo dos valores obtidos, o processo de Descoberta de Conhecimento pode sugerir uma nova iteração

Com a volta à etapa de Min. de Dados, o processo é repetido com valores de parâmetros modificados ou com a utilização de novos algoritmos



# KDD X Mineração de Dados

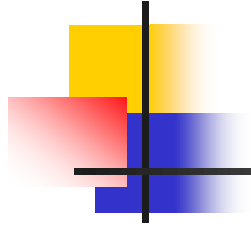
---

Como se viu, a etapa em que efetivamente se dá a descoberta de conhecimento é a da **Mineração de Dados**

Muitas ferramentas originalmente implementadas apenas para a etapa de **Mineração de Dados** atualmente permitem fazer também o **pré** e o **pós-processamento**

Por isso, na prática, vários autores consideram os termos **Descoberta de Conhecimento e Mineração de Dados** como **equivalentes**





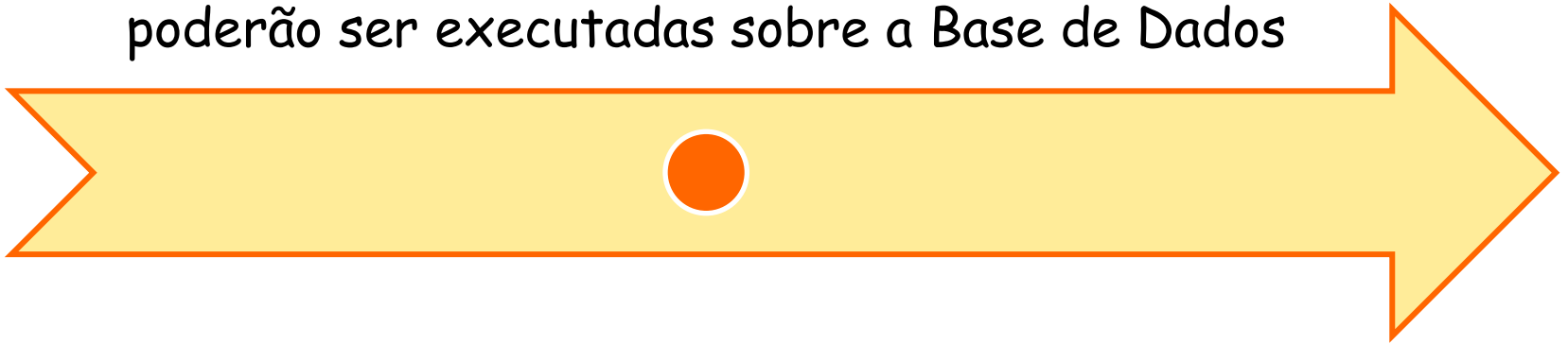
# MINERAÇÃO DE DADOS



# Tipo de Conhecimento Gerado

---

Na **Mineração de Dados (MD)**, dependendo do objetivo a ser atingido, ou seja, do **tipo de conhecimento a ser gerado**, diferentes tarefas poderão ser executadas sobre a Base de Dados





# Principais Tarefas da MD

---

**Associação** - nessa tarefa, partindo de um conjunto de itens o objetivo é aprender a encontrar **Regras de Associação** entre itens que ocorrem simultaneamente

**Classificação** - a partir de exemplos previamente rotulados em **duas ou mais classes**, o objetivo é aprender a classificar um novo exemplo, cuja classe é desconhecida. Se a saída for contínua, a classificação recebe o nome de Regressão

**Clusterização** - também conhecida como Agrupamento, um grupo de registros diversos de uma Base de Dados deve ser segmentado em subgrupos contendo **registros similares**

**Detecção de Anomalias** - o objetivo é detectar desvios de um comportamento considerado normal, e caracterizar uma situação como anormal ou não



# Natureza das Tarefas

As tarefas de Mineração de Dados são de natureza **Descritivas** ou **Preditivas**

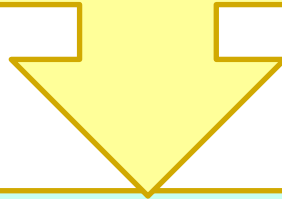
**Descritivas:** detectam padrões de associação  
**Preditivas:** predizem se um item pertence a uma classe ou não



# Implicações Éticas da MD



Como o uso de dados pessoais na **Mineração de Dados** pode afetar a privacidade de pessoas ou discriminar grupos socialmente fragilizados, suas **implicações éticas** têm sido amplamente discutidas na imprensa, na academia, nos meios jurídicos e no mundo corporativo



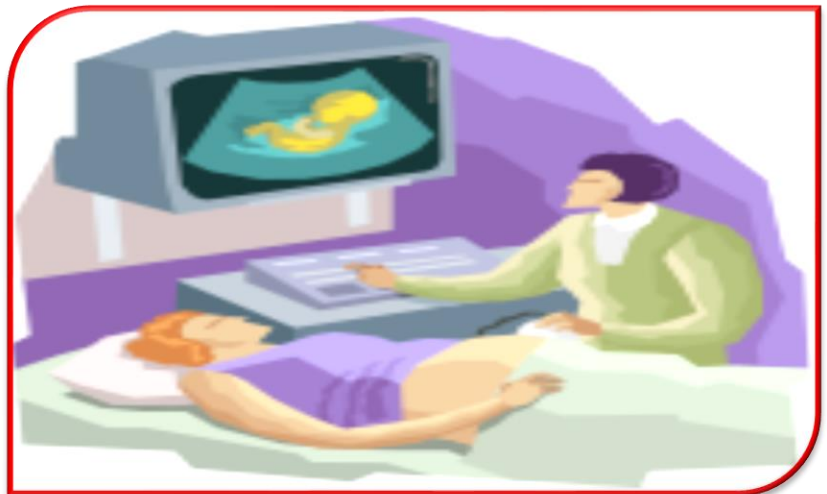
Quando se pensa na política das companhias de seguro, por ex., argumenta-se que **um cliente é geralmente julgado mais pelos atributos do grupo ao qual ele se enquadra** e nem tanto pelas suas próprias características

# Exemplos de Implicações Éticas

Os dados sobre pagamentos feitos com cartão de crédito podem expor as preferências religiosas do dono, assim como seus hábitos de compras de livros podem revelar suas preferências políticas



O CEP de um candidato pode apontar que ele vive em uma região considerada problemática ou num bairro nobre, seus dados médicos podem lhe custar uma vaga numa grande empresa



# Exemplos de Implicações Éticas

Por outro lado, as estatísticas sobre determinada modalidade de crime podem ajudar as autoridades policiais a adotar medidas preventivas



O conhecimento prévio de que algumas doenças parecem estar mais fortemente associadas a uma etnia que a outra auxilia o médico na hora de solicitar exames médicos, mesmo que o paciente não apresente nenhum sintoma





# Considerações Finais

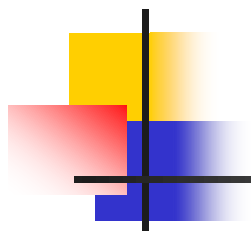
---

Como se vê, não é nada simples traçar uma linha divisória que condene ou justifique o **uso ético** ou legal de dados armazenados

E se os dados tiverem sido coletados sem o conhecimento do usuário, a questão torna-se ainda mais controversa

Por esta razão, e por se tratar de uma tecnologia relativamente jovem, recomenda-se que a **Mineração de Dados seja empregada com cautela** e, se possível, com a anuência prévia das pessoas cujos dados foram coletados





# REFERÊNCIA BIBLIOGRÁFICA



# Referência Bibliográfica

---

- FOROUZAN, B. & MOSHARRAF, F. **Fundamentos da Ciência da Computação**. São Paulo: Cengage Learning, 2011.
- GOLDSCHMIDT, R. & PASSOS, E. **Data Mining: Um Guia Prático**. Rio de Janeiro: Elsevier, 2005.
- HAN, J. & KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2008.
- PADHY, N. P. **Artificial Intelligence and Intelligent Systems**. New Delhi: Oxford University Press, 2010.
- PINHEIRO, C. A. R. **Inteligência Analítica: Mineração de Dados e Descoberta de Conhecimento**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2008.
- REZENDE, S. O. (Organizadora). **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri: Editora Manole Ltda, 2005.
- RUSSEL, S. & NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2004.
- TAN, P.N.; STEINBACH, M. & KUMAR, V. **Introdução ao Data Mining Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009.
- WITTEN, I. H. & FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. Second Edition. Amsterdam: Morgan Kaufmann Publishers, 2005.