

# Aprendizado de Máquina aplicado a Testes

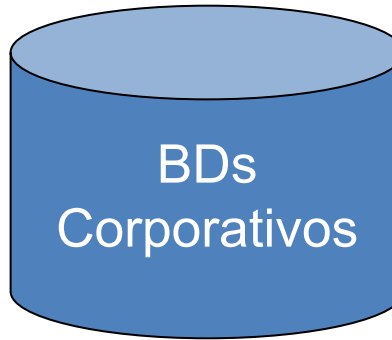
Ricardo Prudêncio

# Big Data

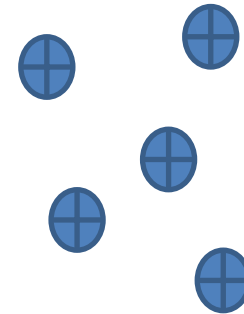
Coisas



BDs  
Corporativos



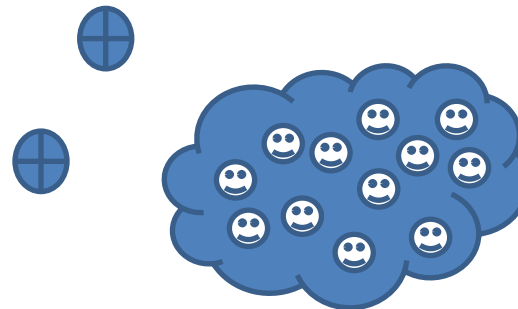
Sensores



Internet,  
WEB

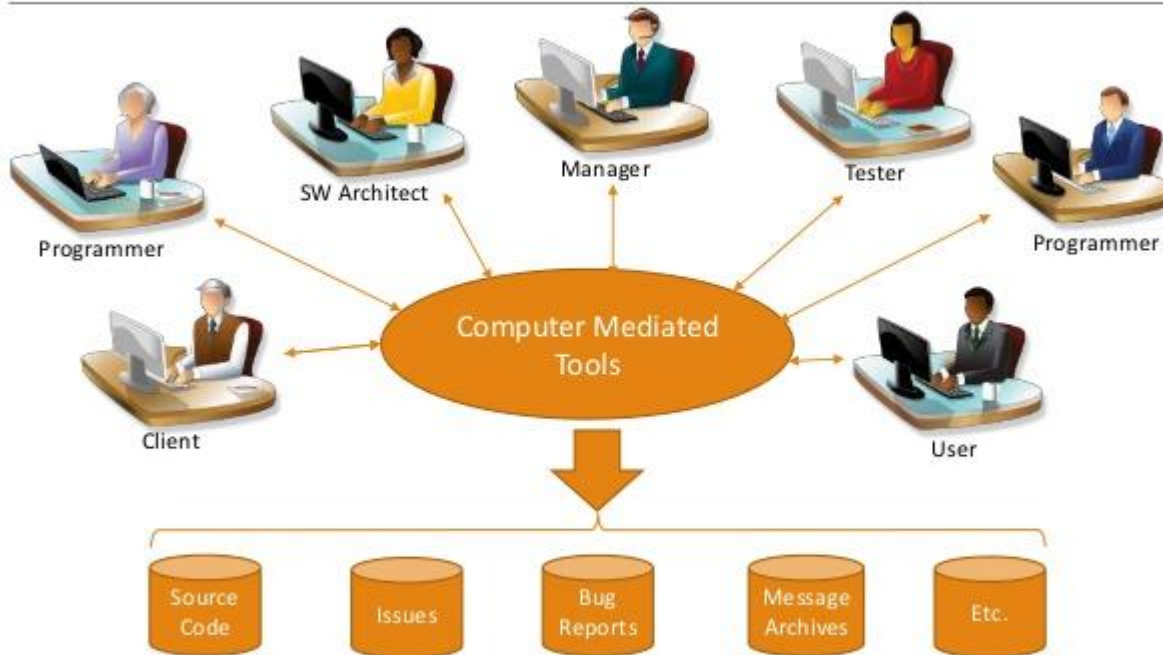


Redes  
Sociais



# Repositórios de Software

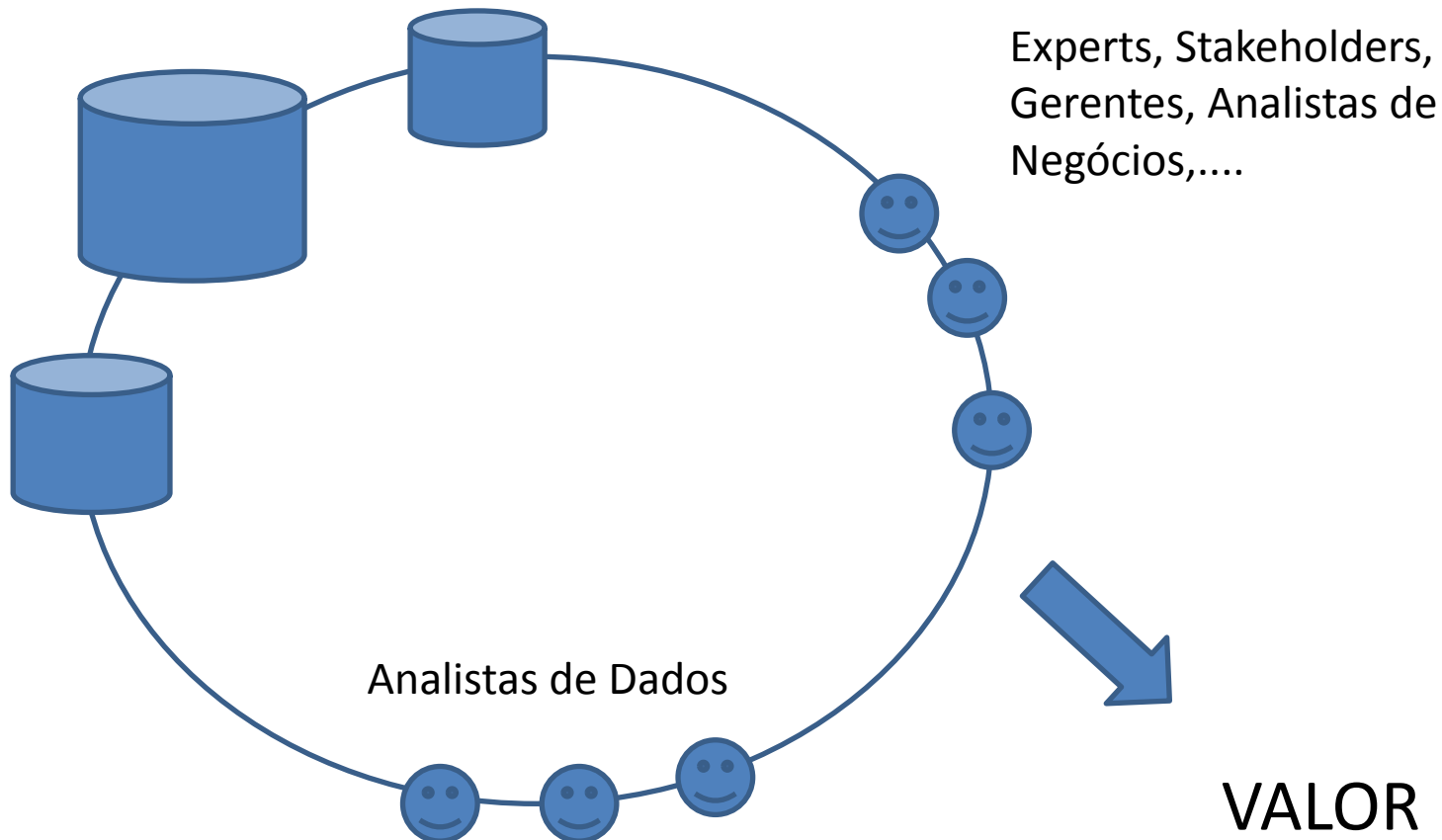
## Software repositories



Current and historical artifacts and interactions are registered in software repositories

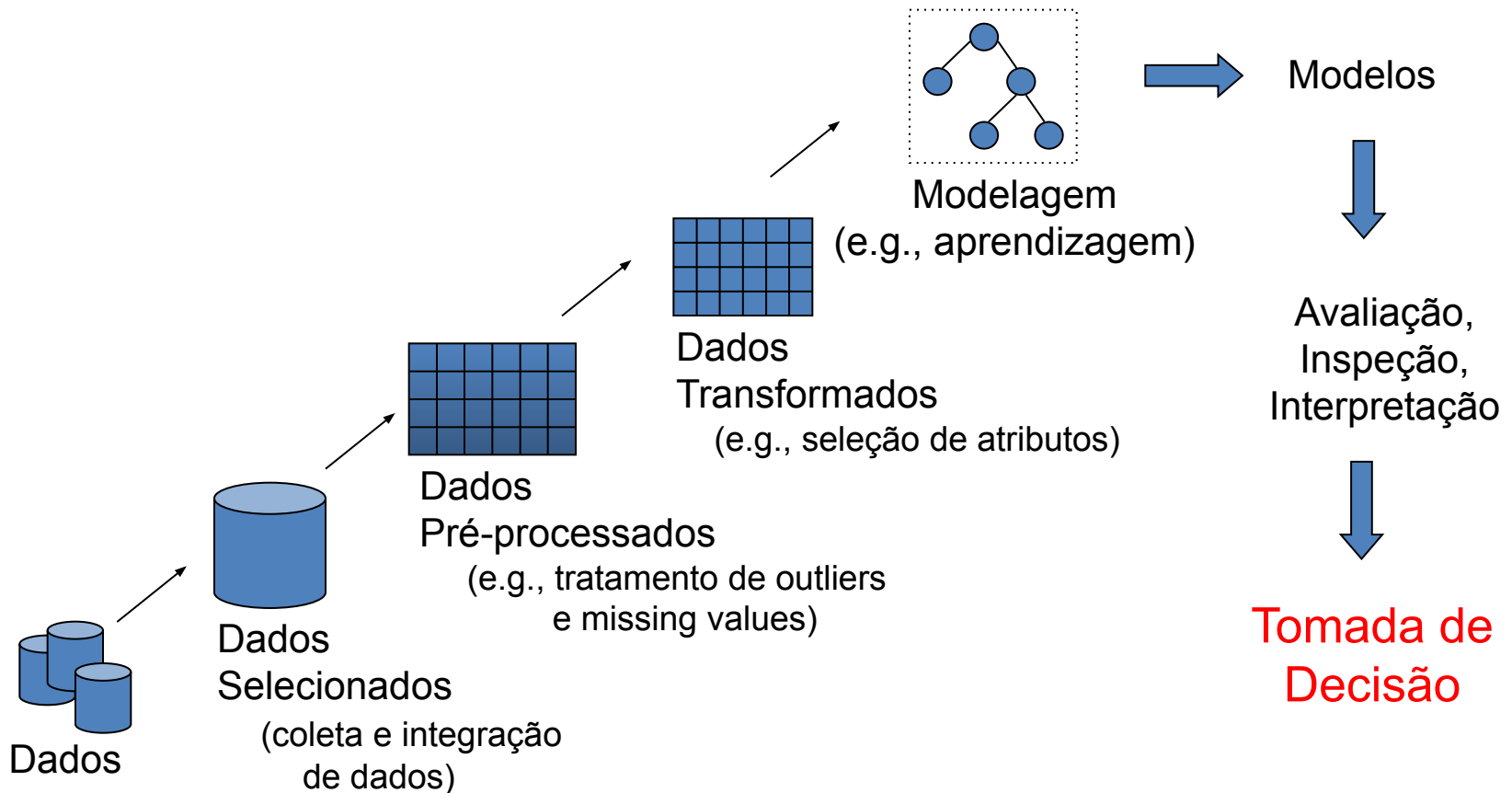
# KDD (Knowledge Discovery in Databases)

Bases de Datos,....



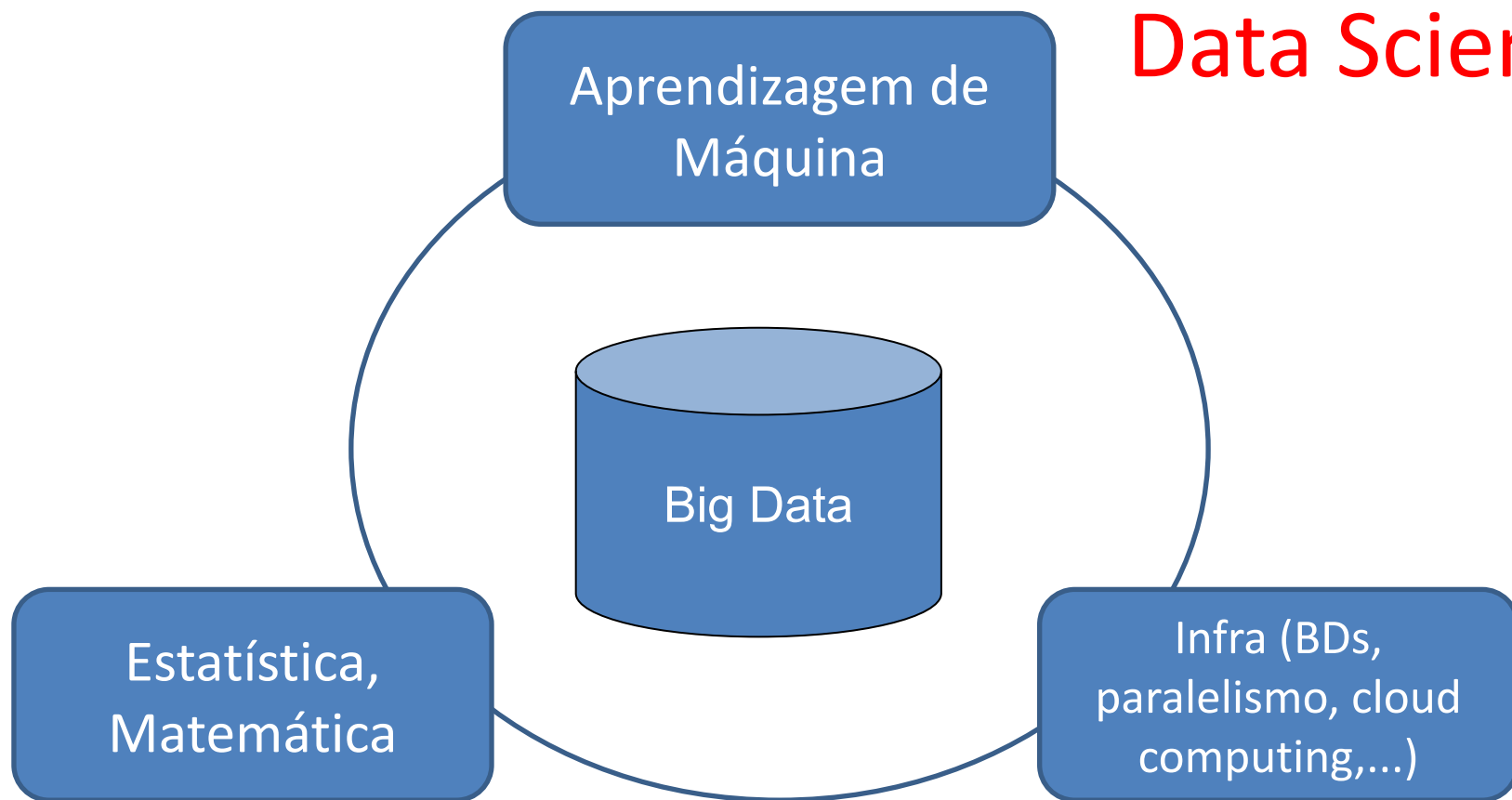
VALOR

# KDD (Knowledge Discovery in Databases)



# KDD, Big Data

Data Science



# Analista, Cientista de Dados

- Definir **objetivos**
- Guiar o processo de **coleta e tratamento** de dados
- Definição de problemas (**tarefas**)
- Construção de soluções (**modelos**)
- **Avaliação e interpretação** dos modelos

# Disciplina

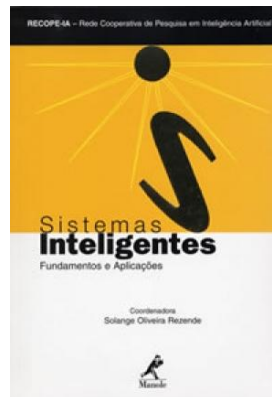
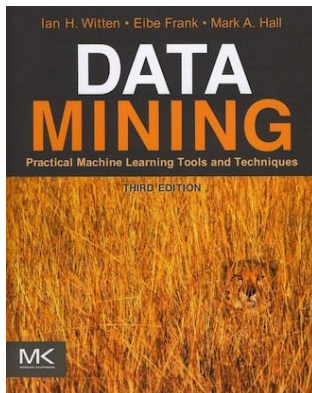
- Foco:
  - Aprendizagem de Máquina
- Conteúdo
  - Análise Exploratória de Dados
  - Problemas
    - Classificação, Regressão e Agrupamento
  - Tópicos Especiais



Introdução e conceitos básicos	02/mai
Algoritmos de classificação (Árvores de Decisão e Random Forests)	03/mai
Algoritmos de classificação (Algoritmos Baseados em Vizinhaça, Regressão Logística, Naive Bayes)	04/mai
Avaliação de algoritmos de classificação	05/mai
Algoritmos de agrupamento / Algoritmos de regressão	08/mai
Pré-processamento de dados e seleção de atributos	09/mai
Acompanhamento de Projetos - Parte 1	10/mai
Acompanhamento de Projetos - Parte 2	11/mai
	t

# Disciplina

- Avaliação
  - Listas de exercício de acompanhamento (peso 5) e uma lista final (peso 5)
- Ferramenta
  - Weka
- Material de Estudo



# Inteligência Artificial

Breve Introdução

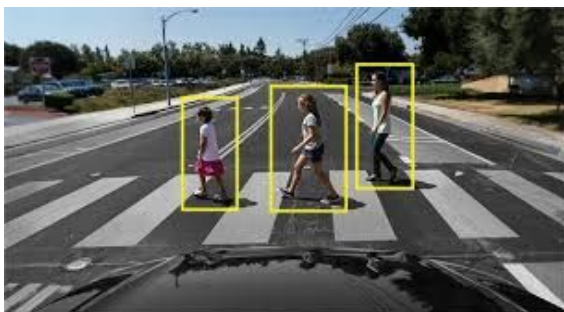
# Exemplos de aplicações de IA

# Robótica

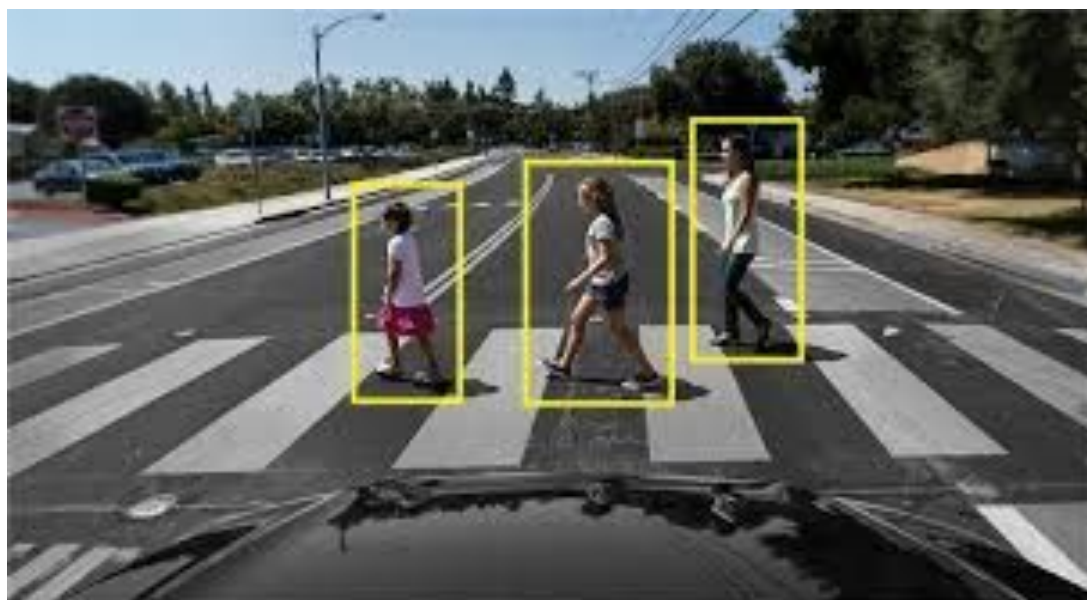
- Como obter manipulação fina e versátil de objetos, navegação segura e eficiente, autonomia?
- E no caso de ambientes dinâmicos e imprevisíveis?



# Carros Autônomos

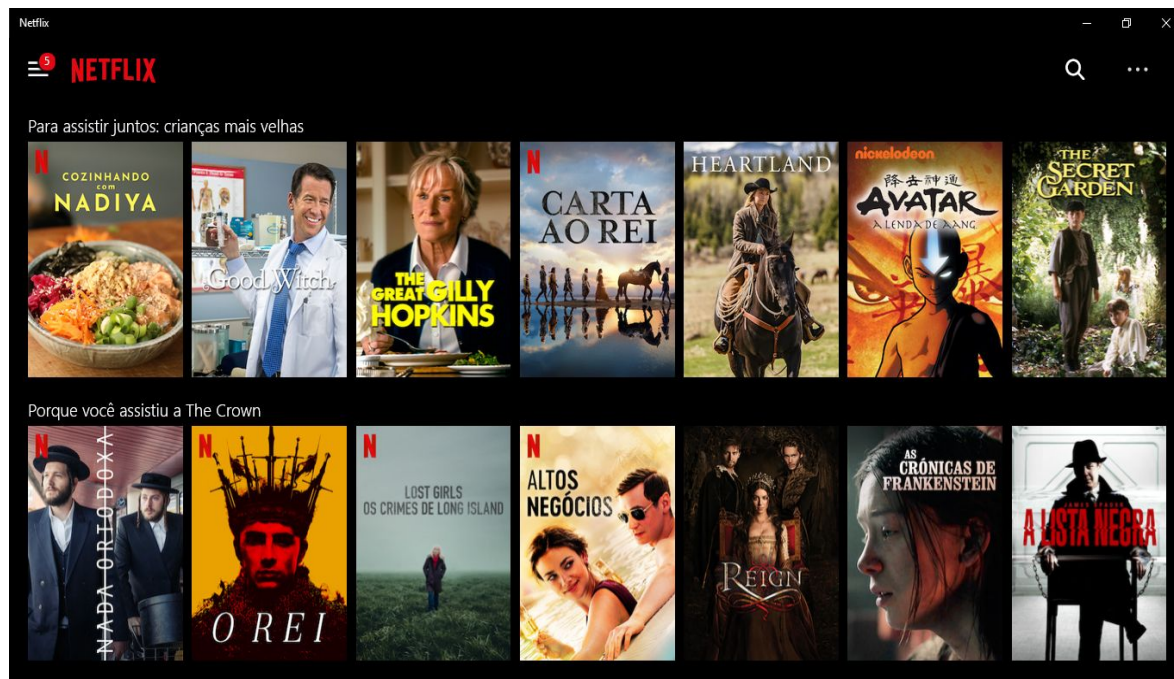


# Reconhecimento de Padrões



# Sistemas de Recomendação

- Como fazer **recomendações personalizadas** de produtos/serviços/experts?
- Como modelar o **perfil** do usuário?





# Interação Humano Computador

- Como dar ao usuário a ajuda de que ele precisa?
- Como adaptar a informação e funcionalidades ao dispositivo e à situação do usuário?
- Como interagir e navegar na web via voz?



# IA em Sistemas computacionais

- Maior **valor agregado** quando embutida em sistemas de computação
  - A cereja do bolo
  - A azeitona da empada



# Voltando às aplicações...

O que esses problemas têm em comum?

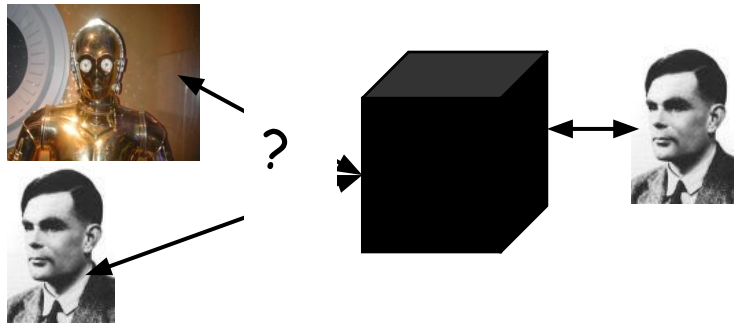
- Grande complexidade
  - número, variedade e natureza das tarefas
- Não há “solução algorítmica” viável
  - mas existe conhecimento (em intenção ou em extensão) sobre o problema
- Requerem modelagem do comportamento de um “ser inteligente”
  - autonomia, aprendizagem, conhecimento, etc.

# Máquinas inteligentes?

- Inteligência Artificial
  - Surgiu na década de 50
  - Objetivo: desenvolver sistemas para realizar tarefas que, no momento
    - ◇ são melhor realizadas por seres humanos que por máquinas, ou
    - ◇ não possuem solução algorítmica viável pela computação convencional

# Teste de Turing

- Máquina que age humanamente:



Teste de Turing envolve:

Processamento de  
Linguagem Natural

Representação de  
Conhecimento

Raciocínio Automático

**Aprendizagem**

Visão Computacional

Robótica

# Aprendizagem de Máquina

- Sub-área de IA que desenvolve sistemas que melhoram seu desempenho com a experiência

# Aprendizagem de Máquina

- Abordagens:
  - Data-driven: algoritmos encontram regularidades em dados
    - Classificação, agrupamento, regressão,...
  - Agent-driven: algoritmos exploram um ambiente para aprender a tomar decisões
    - Aprendizagem por reforço

# Aprendizagem de Máquina

## - Abordagem Data-Driven

- Modelos descritivos
  - Descrevem ou sumarizam dados
  - E.g., Agrupamento
  - Usam dados não rotulados
  - Aprendizado não supervisionado
- Modelos preditivos
  - Realizam previsões sobre os dados
  - E.g., Classificação
  - Usam dados rotulados com uma variável alvo
  - Aprendizado supervisionado



# Modelos Preditivos

## - Classificação

# Classificação

- Associar objetos a uma **categoria** ou classe
  - E.g., diagnóstico de pacientes, classificação risco de um cliente, classificação de documentos,...
- Classificação é feita com base nos **atributos** dos objetos
  - E.g., diagnóstico de um paciente é feito com base nos sintomas observados e exames realizados

# Exemplos

- Diagnóstico médico
  - Qual a chance de reincidência?
  - Onde estão as regiões de risco em uma imagem de tomografia?

# Exemplos

- Classificação de imagens
  - Você é você mesmo? :)
  - Onde estão as anomalias nas imagens?
  - Essa imagem tem conteúdo sensível?

# Exemplos

- Detecção de anomalias
  - Que transações de crédito são fraudes?
  - Quais equipamentos vão falhar?

# Exemplos

- Categorização de usuários
  - Que usuários do Instagram se interessam por roupas? Ou automóveis? Ou perfumaria?
  - Qual o viés político de um usuário no Twitter?

# Exemplos

- Análise de sentimentos
  - Esse Twitter expressão raiva, alegria, angústia?
  - Essa pessoa está feliz?

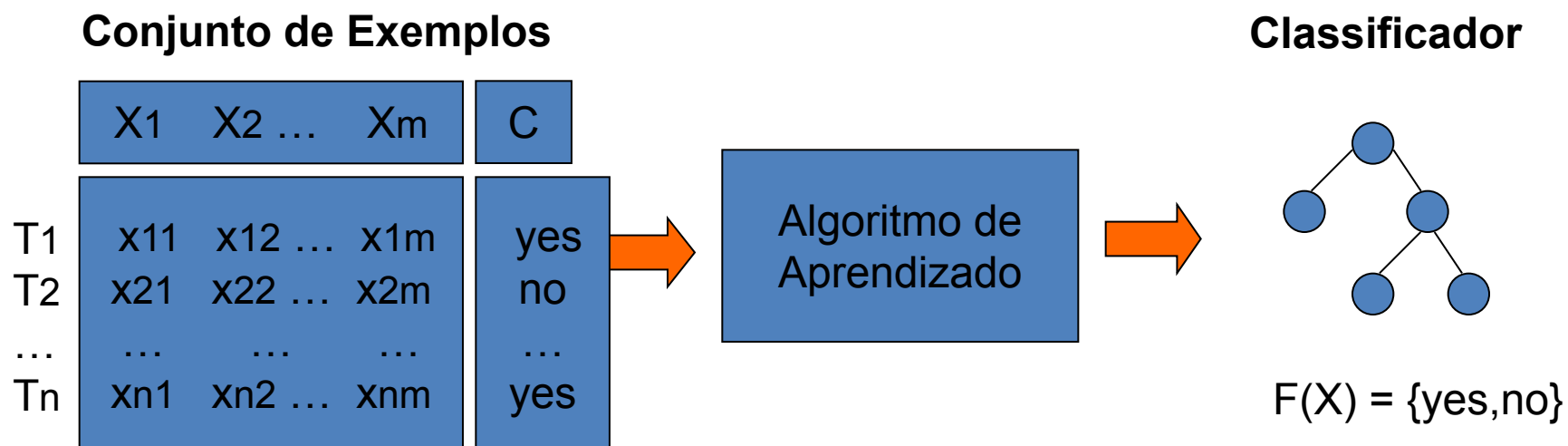
# Exemplos

- Funções de score
  - Quem será um bom pagador?
  - Qual o risco de sonegação?
  - Quem contratar?



# Classificação com AM

- Algoritmo de **aprendizagem supervisionada** adquire conhecimento a partir de um conjunto de exemplos

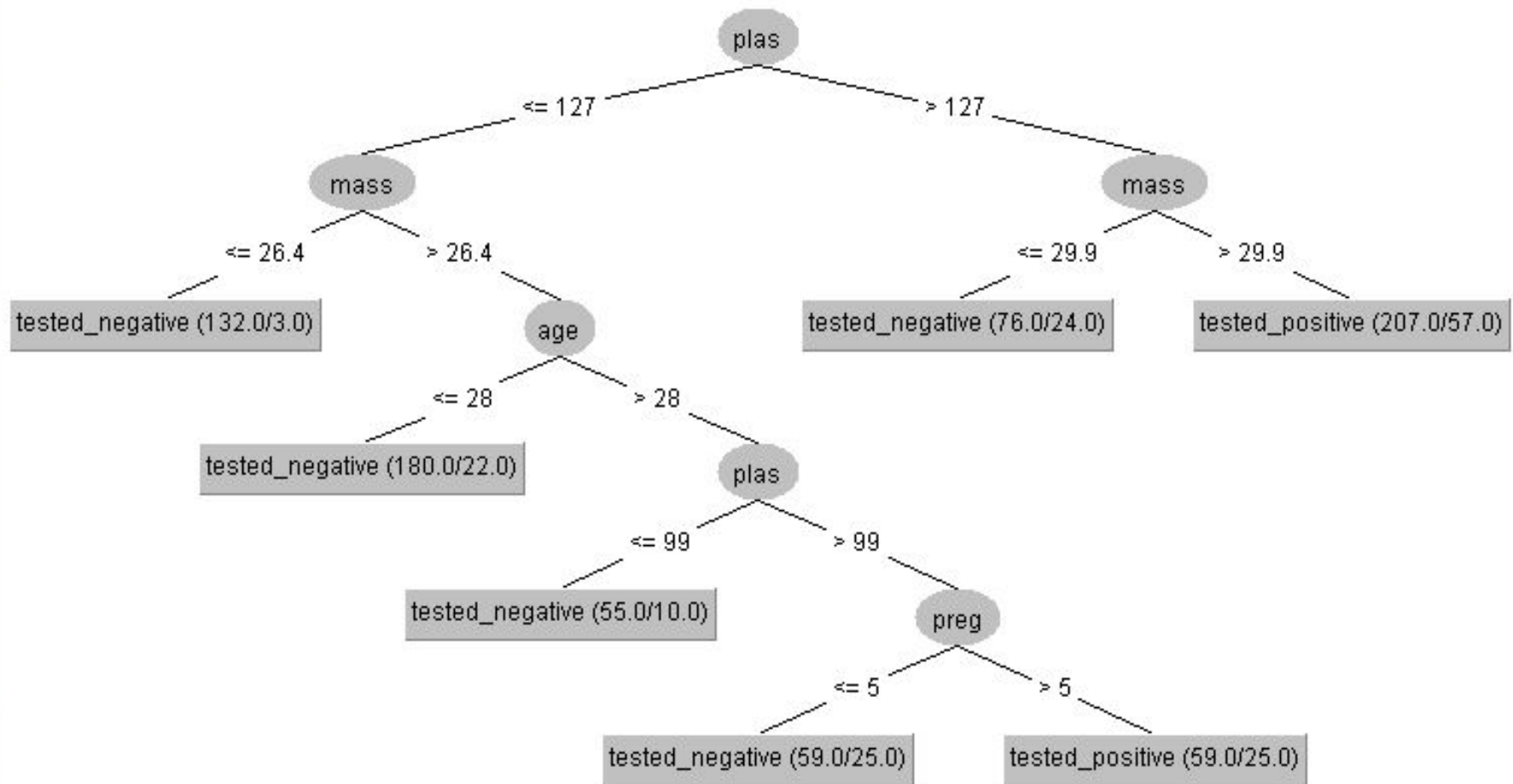


# Exemplo

## Conjunto de Dados - Diabetes

1: preg Numeric	2: plas Numeric	3: pres Numeric	4: mass Numeric	5: age Numeric	6: class Nominal
6.0	148.0	72.0	33.6	50.0	tested_positive
1.0	85.0	66.0	26.6	31.0	tested_negative
8.0	183.0	64.0	23.3	32.0	tested_positive
1.0	89.0	66.0	28.1	21.0	tested_negative
0.0	137.0	40.0	43.1	33.0	tested_positive
5.0	116.0	74.0	25.6	30.0	tested_negative
3.0	78.0	50.0	31.0	26.0	tested_positive
10.0	115.0	0.0	35.3	29.0	tested_negative
2.0	197.0	70.0	30.5	53.0	tested_positive
8.0	125.0	96.0	0.0	54.0	tested_positive
4.0	110.0	92.0	37.6	30.0	tested_negative
10.0	168.0	74.0	38.0	34.0	tested_positive
10.0	139.0	80.0	27.1	57.0	tested_negative
1.0	189.0	60.0	30.1	59.0	tested_positive
5.0	166.0	72.0	25.8	51.0	tested_positive
7.0	100.0	0.0	30.0	32.0	tested_positive
0.0	118.0	84.0	45.8	31.0	tested_positive
7.0	107.0	74.0	29.6	31.0	tested_positive
1.0	103.0	30.0	43.3	33.0	tested_negative
1.0	115.0	70.0	34.6	32.0	tested_positive
3.0	126.0	88.0	39.3	27.0	tested_negative
8.0	99.0	84.0	35.4	50.0	tested_negative
7.0	196.0	90.0	39.8	41.0	tested_positive
9.0	119.0	80.0	29.0	29.0	tested_positive
11.0	143.0	94.0	36.6	51.0	tested_positive
10.0	125.0	70.0	31.1	41.0	tested_positive
7.0	147.0	76.0	39.4	43.0	tested_positive
1.0	97.0	66.0	23.2	22.0	tested_negative
13.0	145.0	82.0	22.2	57.0	tested_negative

# Exemplo - Modelo de Árvore de Decisão



# Exemplo - Modelo de Regressão Logística

Class	
Variable	tested_positive
=====	
preg	0.1188
plas	0.0338
pres	- 0.0135
mass	0.091
age	0.0171

$$P(Y = 1 | x_1, \dots, x_p) = \frac{1}{1 + \exp(-(\beta + \alpha_1 x_1 + \dots + \alpha_p x_p))}$$

# Classificação - Definições

- **Exemplo** (ou instância)
  - Tupla com atributos que descrevem um objeto de interesse + classe do exemplo
    - E.g., dados de um paciente + doença
    - E.g., medidas de complexidade de software + {bug ou não bug}
- Atributos **Preditores**
  - Característica de um exemplo usada para classificação
- Atributo **Alvo**
  - Problemas de classificação binários ou multi-class

# Classificação - Definições

- Conjunto de Treinamento
  - Coletado da base de dados e **etiquetado (rotulado)** geralmente por um humano
  - Usado para construir um classificador
- Conjunto de **Teste**
  - Conjunto usado para **avaliar** a qualidade do classificador gerado
- Classificador (**Modelo**)
  - Resultado retornado pelo indutor (aproxima a função real de classificação)

## Ver WEKA

- Conjunto Iris
- Conjunto Heart-Statlog
- Conjunto Change-Metrics

# Aprendizado de Máquina Supervisionado

- Categorias de algoritmos
  - Árvores de decisão e regras
  - Aprendizado lazy
  - Aprendizado Bayesiano
  - Redes neurais
  - Métodos de Kernel,
  - Etc...
- Qual o **melhor** algoritmo?



# Referências

- T. Mitchell, *Machine Learning*, Cap. 3, 1997.
- I. Witten, E. Frank, 2000. *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations*.
- M. Monard, J. Baranauskas, Indução de Regras e Árvores de Decisão, *Sistemas Inteligentes*, Cap. 5, 2005.
- J. R. Quinlan, Induction of Decision Trees, *Machine Learning*, Vol.1, N.1, 1986.