



# ANALISTA DE DADOS

## | Nov22

# SOBRE MIM



## SAMUEL TREMÉA

- 25 anos;
- Formado em Estatística;
- Consultor trainee;
- Trabalha na empresa há 1 ano e 6 meses;
- Hobbies: cantar e escrever.

# INTRODUÇÃO A ALGORITMOS

# ESTATÍSTICA DESCRITIVA

## Correlação - Variáveis

### VARIÁVEL

Característica de interesse que é medida em cada elemento da amostra ou população;

Seus valores variam de elemento para elemento e podem ter valores numéricos ou não numéricos.

### QUALITATIVA

Representa uma característica da qualidade (ou atributo) associado ao item pesquisado;

Podem ser divididas em dois tipos: qualitativa nominal; qualitativa ordinal.

### QUANTITATIVA

São as características que podem ser medidas. Neste caso, a característica observada assume valores numéricos que podem ser classificados em "discretos" ou "contínuos".

# ESTATÍSTICA DESCRITIVA

## Correlação - Variáveis

### CATÉGÓRICAS

#### NOMINAL

Não existe ordenação dentre as categorias.

#### ORDINAL

Existe uma ordenação entre as categorias.

### NÚMERICAS

#### DISCRETA

Os possíveis valores formam um conjunto finito ou enumerável de números, são variáveis de contagem.

#### CONTÍNUA

Os possíveis valores estão dentro de um intervalo, aberto ou fechado, dos números reais.

### MODELAGEM

#### INDEPENDENTE/PREDITORA

Influencia, determina ou afeta outra variável.

#### DEPENDENTE/RESPOSTA

Valores a serem explicados, descobertos, observados, medidos.

#### BINÁRIA

Qualitativa, usada para indicar prese./ausência; sucesso/fracasso; Assume apenas valores 0 e 1.

#### CONTAGEM

Quantitativa, guarda valores de contagem.

# ALGORITMOS ESTATÍSTICOS

## Diferenças entre IA e ML

### INTELIGÊNCIA ARTIFICIAL

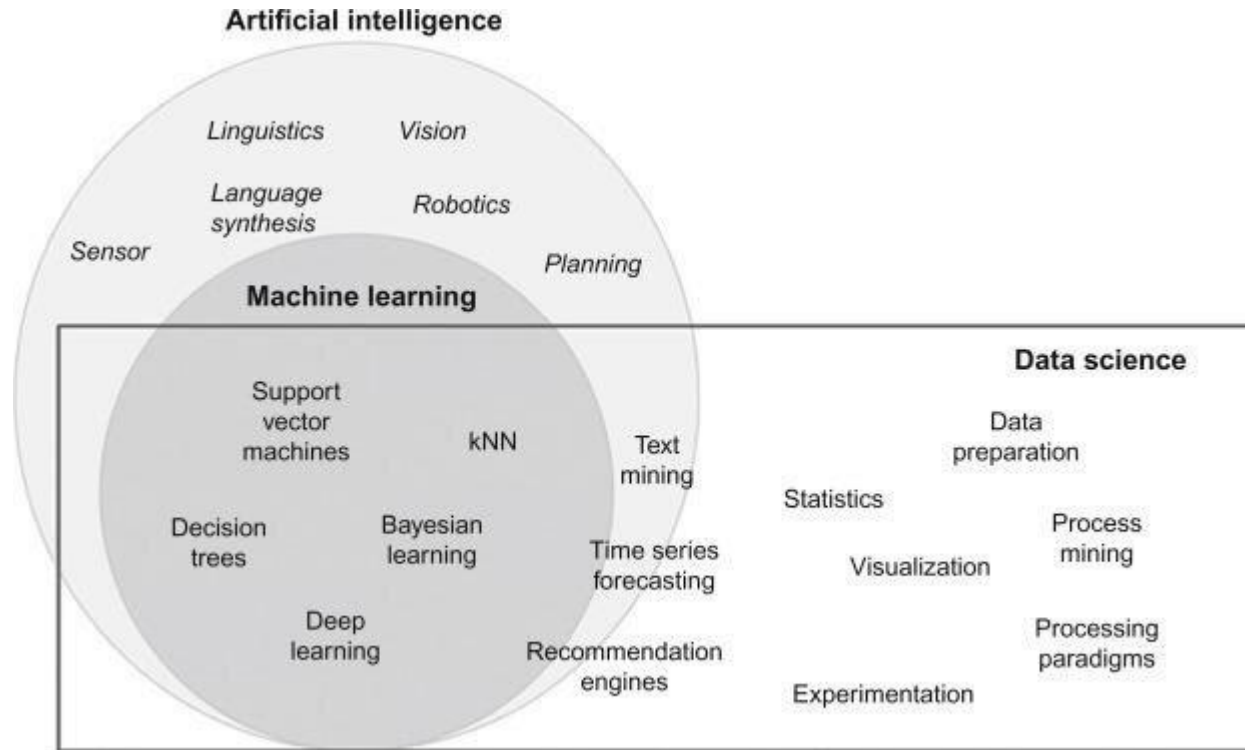
É uma tecnologia que busca **criar sistemas inteligentes que podem simular a inteligência humana**, com o objetivo de resolver problemas complexos.

### MACHINE LEARNING

É uma das maneiras utilizadas pela IA para adquirir uma forma particular de inteligência humana, com foco em ser capaz de **prever resultados quando novos dados forem apresentados** ou apenas **descobrir os padrões ocultos em dados não rotulados**.

# ALGORITMOS ESTATÍSTICOS

## IA, Machine Learning e Ciência de Dados



# ALGORITMOS ESTATÍSTICOS

## Conceitos Importantes de Machine Learning

**ALGORITMO:** conjunto de regras e técnicas estatísticas usadas para aprender padrões de dados e extrair informações significativas deles. Também pode ser interpretado como a lógica por trás de um modelo. **Exemplo: algoritmo de regressão linear.**

**MODELO:** é a expressão/equação gerada através de um algoritmo de machine learning. Um algoritmo mapeia todas as decisões que um modelo deve tomar com base na entrada fornecida.





# ALGORITMOS ESTATÍSTICOS

## Conceitos Importantes de Machine Learning

**DADOS DE ENTRADA:** é o banco de dados original que alimentará o modelo de machine learning.

**DADOS DE TREINAMENTO:** é um subconjunto dos dados de entrada, utilizado para treinar o modelo de machine learning.

**DADOS DE TESTE:** é outro subconjunto dos dados de entrada. Depois que o modelo é treinado, utiliza-se estes dados para avaliar a precisão com que se pode prever um resultado.



# ALGORITMOS ESTATÍSTICOS

## Exemplos práticos

**PREVISÃO DE VENDAS:** Algoritmos podem ser utilizados para realizar uma vasta gama de previsões de diferentes naturezas. Um dos exemplos mais clássicos de suas aplicações é a previsão de vendas de um determinado produto.

Ano	Publicidade	Vendas
1990	2,8	4,3
1991	3,5	5,2
1992	4,6	6,1
1993	6,2	8,3
1994	7,9	9,8



Vendas                      Publicidade

↑                                      ↑

$$Y = \beta_0 + \beta_1 x$$

↓                                      ↓

Valores que queremos descobrir



# ATIVIDADE PRÁTICA

# Descrição da Atividade

- Cada grupo receberá três situações que simularão um problema/estratégia de negócio e o objetivo da atividade é chegar, em termos práticos (não computacionais), em uma solução para a situação utilizando o(s) algoritmo(s) adequado(s).



# Situação 1



- Com dados coletados nos últimos 5 anos a respeito de vendas de bicicletas e automóveis, uma empresa de automobilismo tem interesse em prever o faturamento líquido esperado para o ano de 2023 neste setor.



# Situação 2



- Com base no banco de dados de todos os seus clientes, uma empresa de seguros tem interesse em identificar qual é a porcentagem de clientes que estão vulneráveis/propensos a sofrerem uma fraude.



# Situação 3



- Uma rede de farmácias quer analisar o comportamento de vendas de um determinado remédio para gripe ao longo dos últimos dez anos, buscando identificar algum padrão nas compras neste período de tempo.



# Descrição da Atividade

Na resolução, o grupo deverá responder:

- a)** Por que o(s) algoritmo(s) selecionado(s) pode(m) resolver este problema?
- b)** Quais variáveis preditoras seriam interessantes de se coletar para este problema? Por quê?
- c)** Qual/quais seriam as variáveis-resposta para esse problema? Por quê?
- d)** Qual seria a saída do modelo?
- e)** Quais benefícios para a empresa são esperados com a implementação deste(s) algoritmo(s)?



# INTRODUÇÃO A R

# INTRODUÇÃO A R

## R - O que é?

- R é uma linguagem de programação estatística e gráfica especializada em manipulação, análise e visualização de dados;



# INTRODUÇÃO A R

## R - Por que usaremos?

- Alto poder analítico;
- Simples e gratuito;
- Facilmente integrável;
- Especializado em análises estatísticas.



# INTRODUÇÃO A R

## RStudio - O que é?

- RStudio é uma IDE, (ambiente de desenvolvimento integrado), para o R.



# INTRODUÇÃO A R

## RStudio - Por que usaremos?

- Extremamente amigável;
- Extremamente visual;
- Ambiente intuitivo;

