



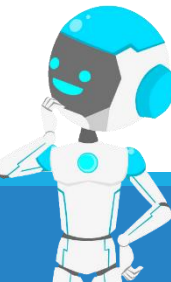
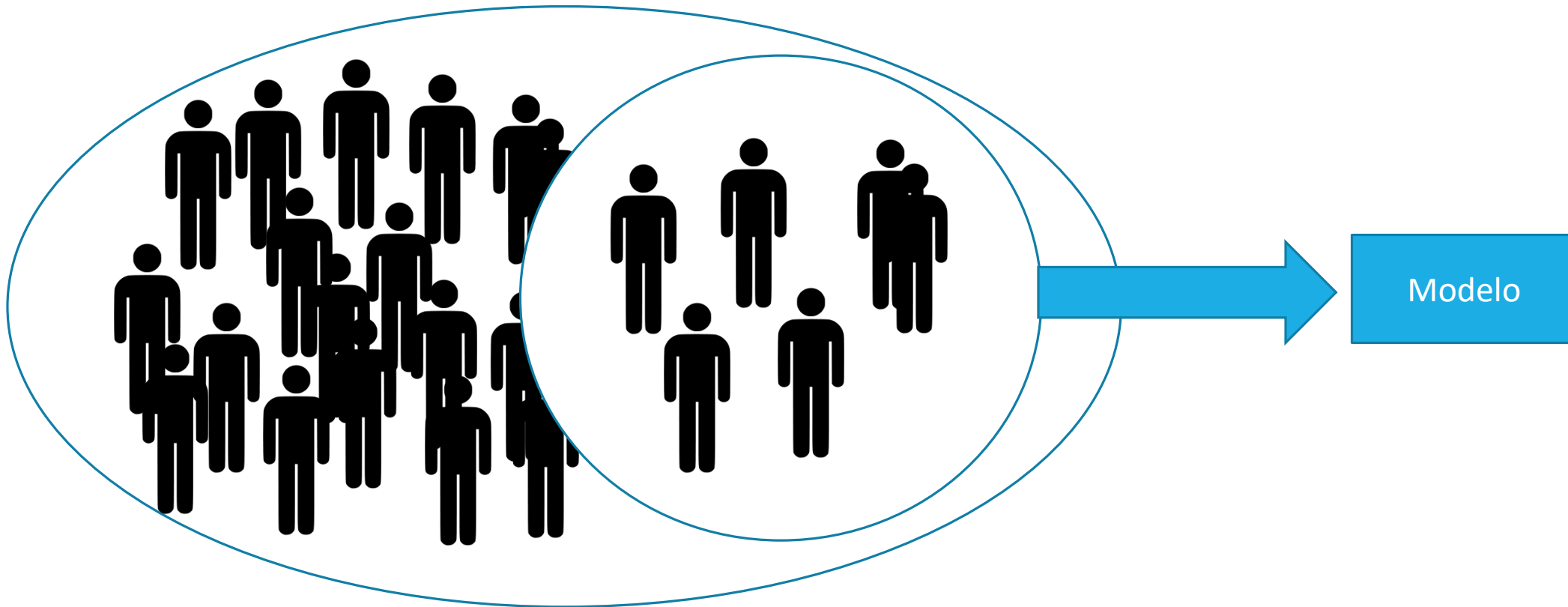
# FORMAÇÃO INTELIGÊNCIA ARTIFICIAL E MACHINE LEARNING

---

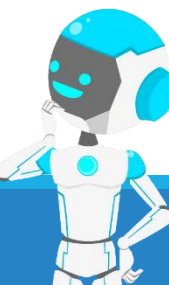
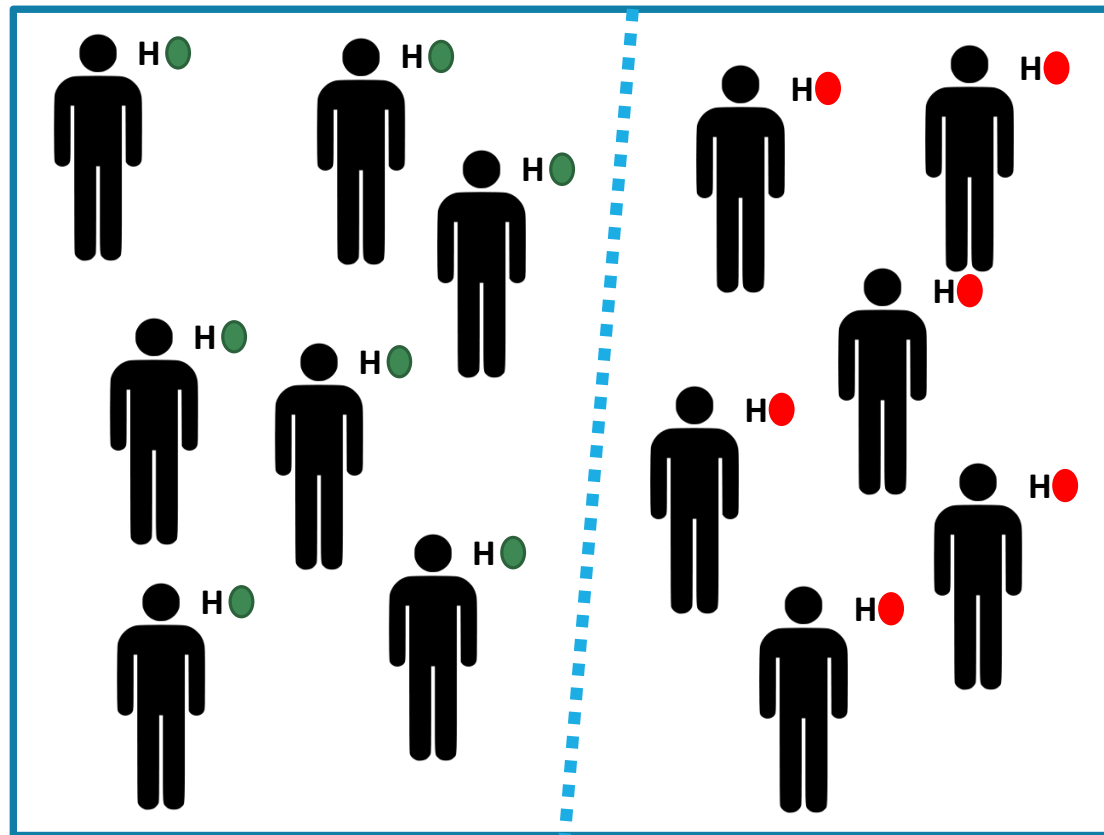
MACHINE LEARNING – TÓPICOS AVANÇADOS  
MALDIÇÃO DA DIMENSIONALIDADE E SELEÇÃO DE ATRIBUTOS

Prof. Fernando Amaral – Todos os Direitos Reservados

# Modelo: representação parcial da realidade

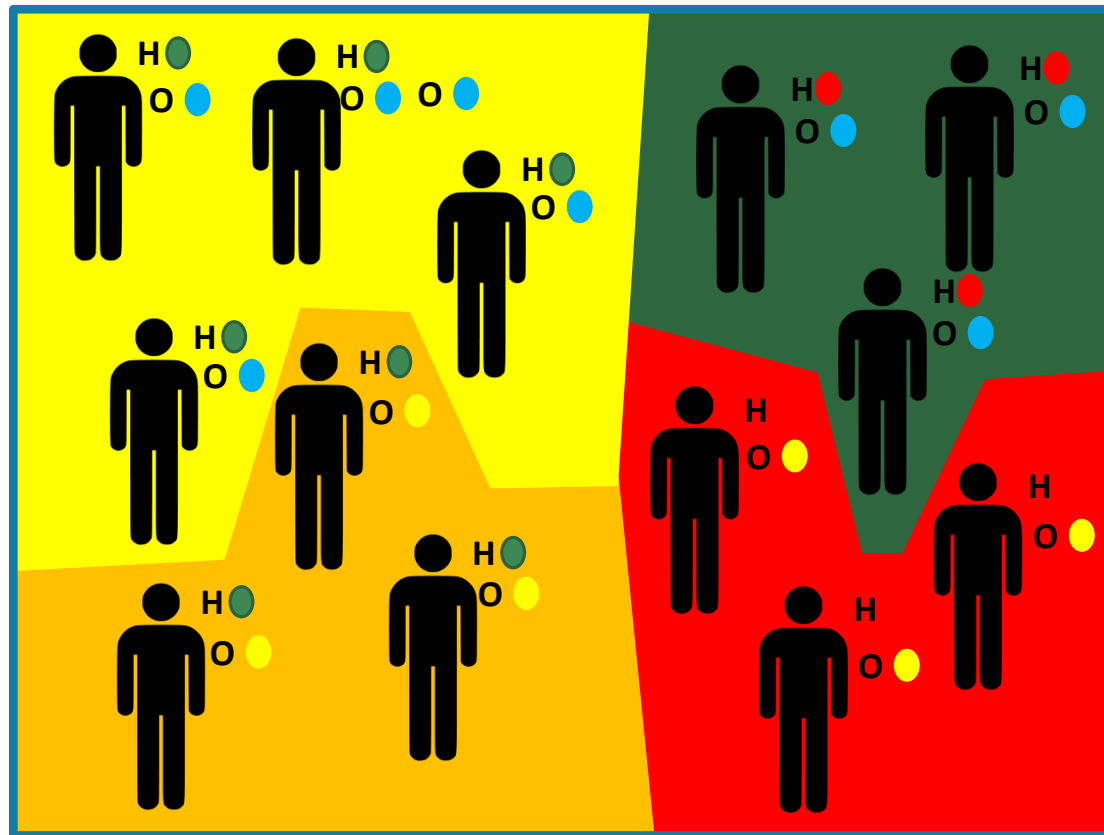


# Um atributo

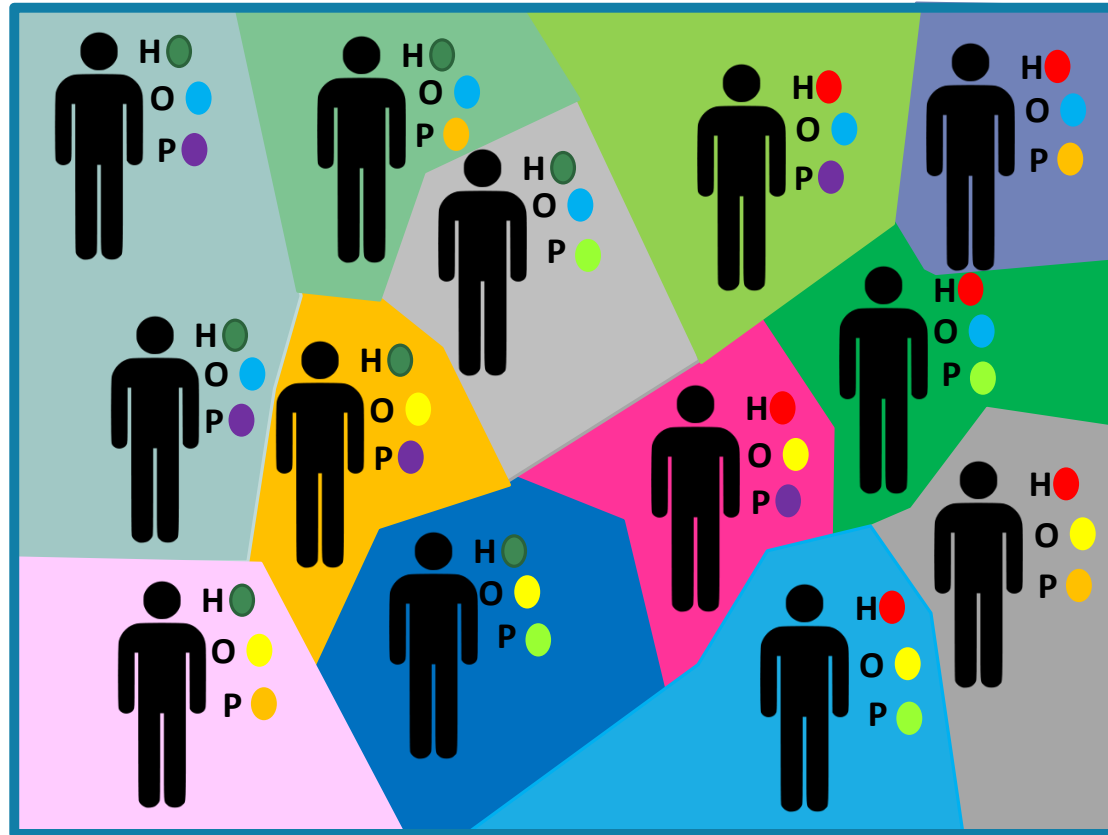


# Duas características

---



# N Características



# Maldição da Dimensionalidade

---

- A inclusão de muitas características do modelo deteriora sua performance, tornando o modelo super ajustado.



# Quantos? Quais?

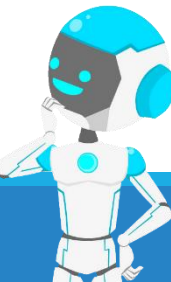
---



# Seleção de Atributos

---

- Qual é o subconjunto de atributos que torna o modelo mais genérico?
  - Atributos generalistas vs atributos específicos
  - Atributos não correlacionados
- Algumas técnicas
  - Força Bruta
  - Testes Estatísticos: Anova, Chi-Square

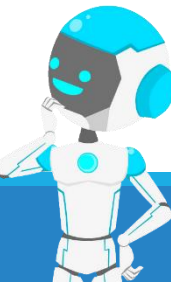




# Internet Advertisements Data Set

---

- Atributos com características de Imagens
- Instâncias: 3279
- Atributos: 1558
- Tarefa: prever se a imagem é um anuncio ou não
  
- Data set original sem nomes de colunas
- Classe: V1559: “ad” ou “nonad”



# Opções

---

- Pacote Boruta

- Força bruta
- Muito eficiente
- Pode levar muitas horas de processamento!

- `modelo = Boruta(V1559 ~. , data=anunciostreino, doTrace=2)`

- Pacote Fselector

- Várias funções de seleção de atributos
- `chi.squared`

