

Introdução à Ciência dos Dados

Gabriel Pacheco

Bacharel em Sistemas de Informação (UFMG)

Mestrando em Ciência da Computação (UFMG)

Trabalha em Digicade desde 2016





Cronograma

- Conceitos e definições
- Aprendizado de Máquina
- Exemplo prático
- Ferramentas



O que é Ciência dos Dados?

A ciência de dados é um campo interdisciplinar que utiliza métodos, processos, algoritmos e sistemas científicos para extrair valor dos dados. Os cientistas de dados combinam uma série de habilidades, incluindo estatísticas, ciência da computação e conhecimento comercial, para analisar dados coletados da web, smartphones, clientes, sensores e outras fontes.



O que é Inteligência Artificial?

Inteligência Artificial (IA) é um ramo da ciência da computação que se propõe a elaborar dispositivos que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas, enfim, a capacidade de ser inteligente.



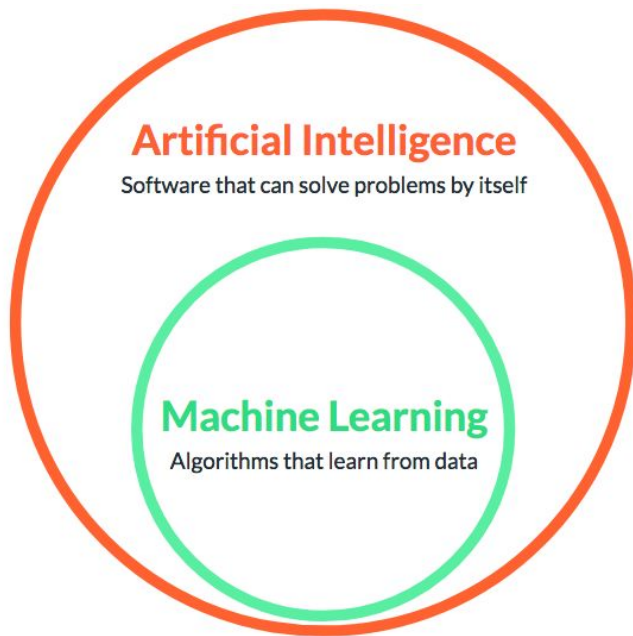
O que é Aprendizado de Máquina?

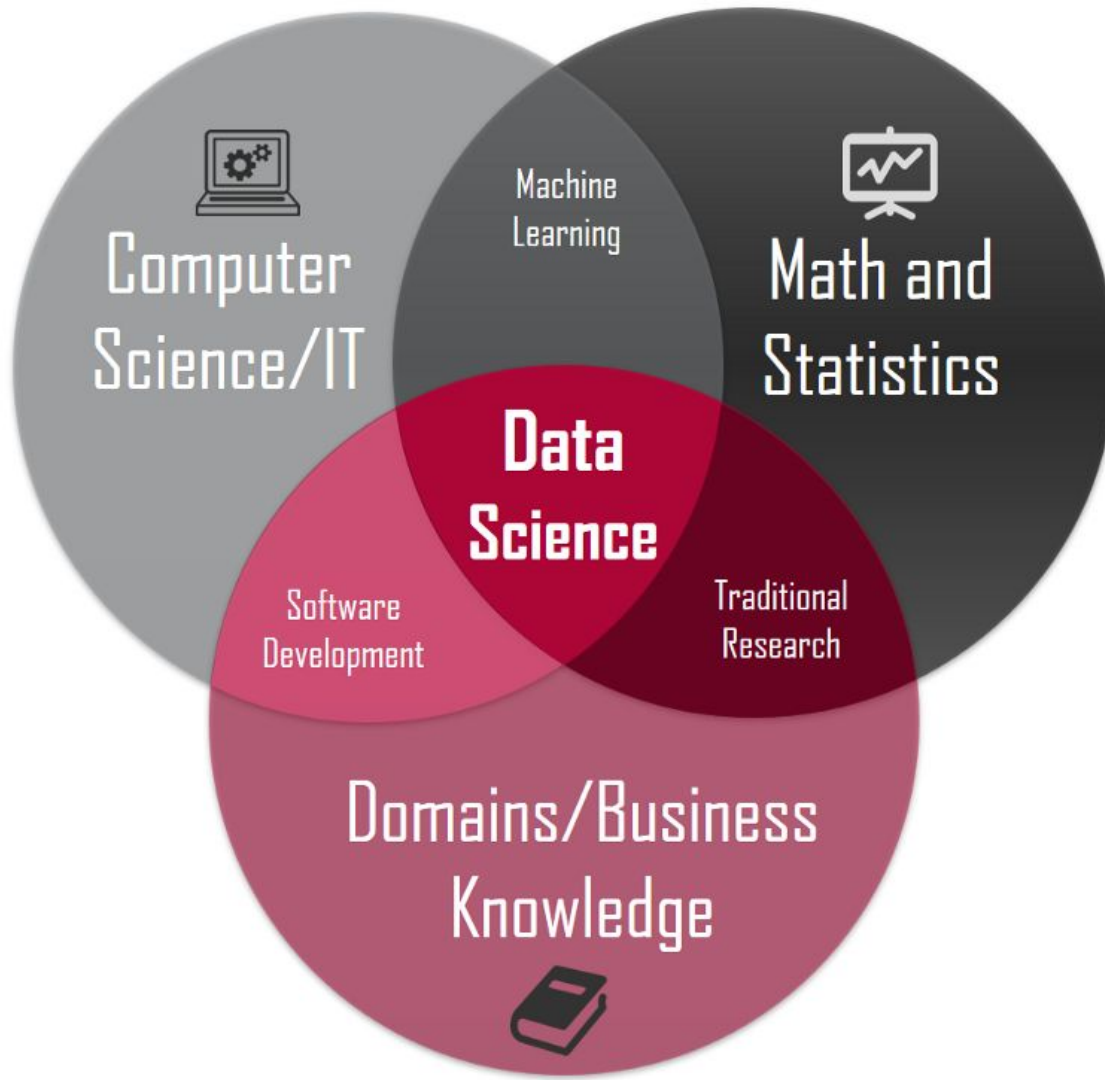
A tecnologia de aprendizagem de máquina trabalha com a ideia de que as máquinas podem aprender sozinhas ao terem acesso a grandes volumes de dados.

A definição mais simples é que as máquinas podem detectar padrões e criar conexões entre dados, por meio de Big Data e algoritmos sofisticados, para aprenderem sozinhas a executar uma tarefa.



Inteligência Artificial e Aprendizado de Máquina





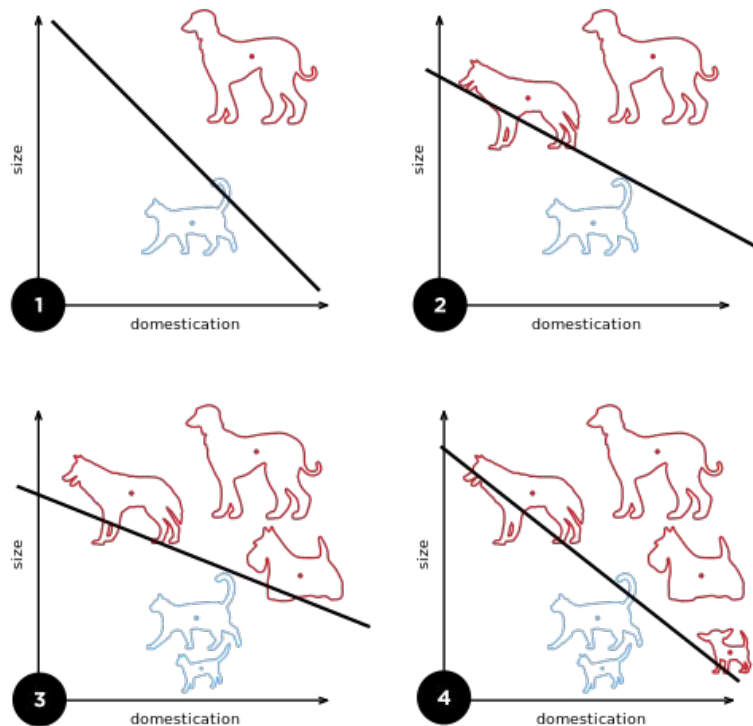


Aprendizado de Máquina

- Classificação
- Associação
- Agrupamento
- Regressão

Classificação

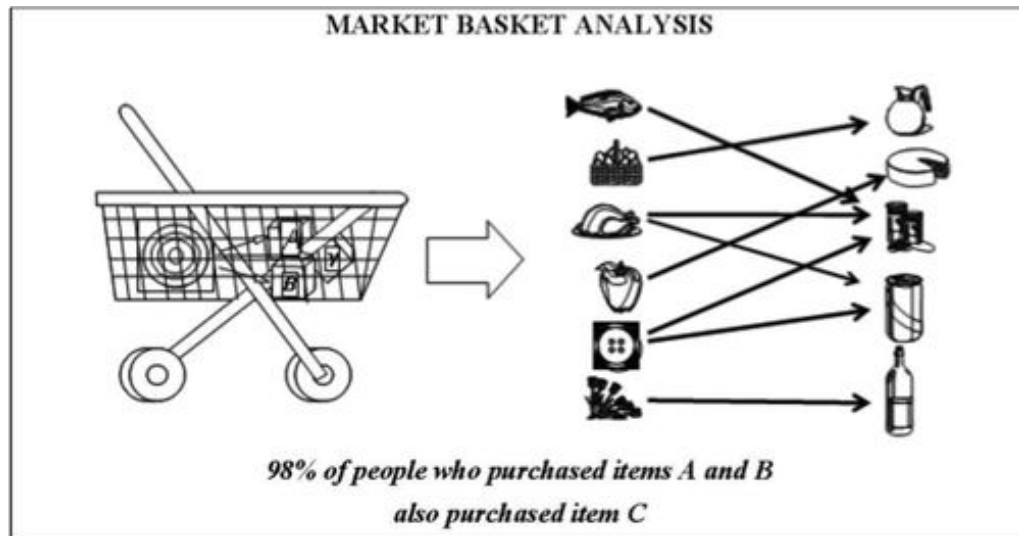
- Detecção de fraudes
- Crédito bancário
- Detecção de Spam
- Análise de Sentimentos
- Reconhecimento de imagem
- Reconhecimento de vídeo
- Análise de ações do mercado financeiro





Associação

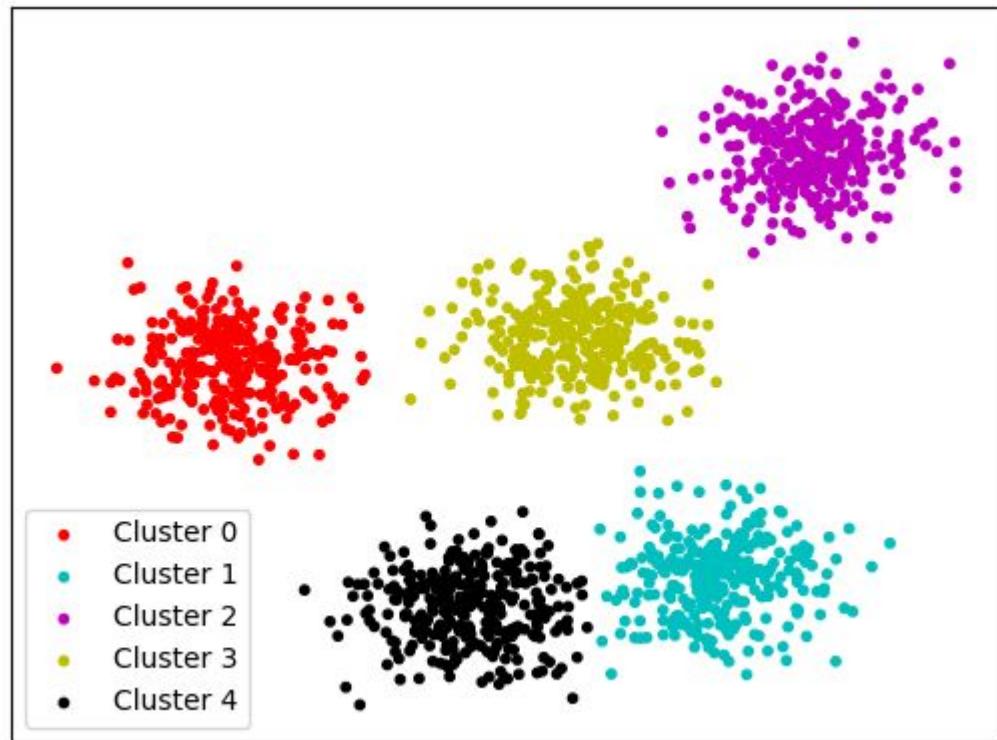
- Recomendação de produtos
 - Filmes
 - Música
 - E-commerce
- Bioinformática
- Árvores...





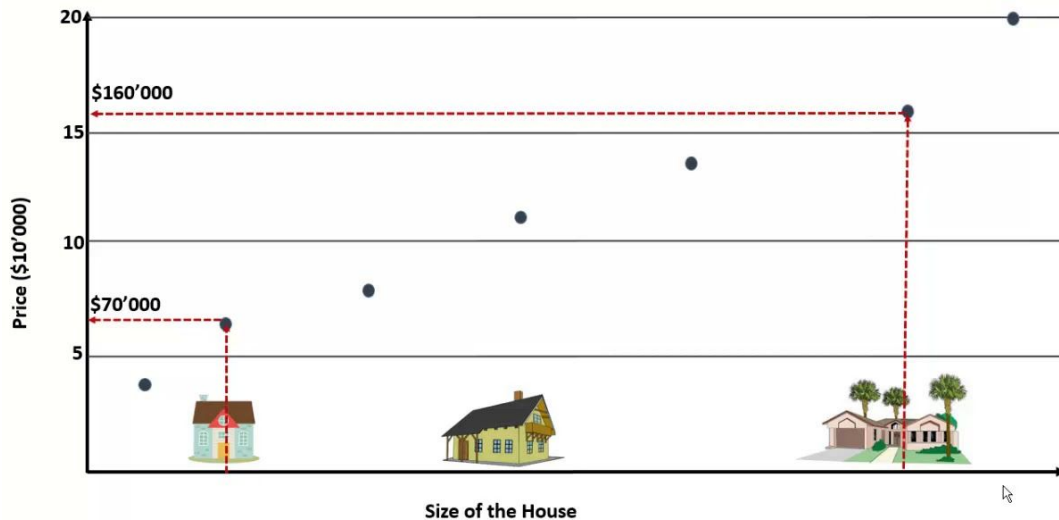
Agrupamento

- Reconhecimento de padrões
- Recuperação de informação
- Detecção de comunidades



Regressão

- Predição de preços
- Predição de demanda*
- Predição de idade





Exemplo prático: Regressão Linear

- Cenários?
 - Atributos
 - Dificuldades
 - Validação



Ferramentas e Links úteis

- Ferramentas e bibliotecas para programação
 - [Jupyter](#)
 - [Scikit-Learn](#)
 - [Pandas](#)
- Ferramentas “drag-and-drop”
 - [Lemonade](#)
 - [Knime](#)
 - [Weka](#)
 - [Power BI](#)
- Base de dados e competições para treinar
 - [Kaggle](#)



Outros tópicos

- Treino, teste e validações
- Limpeza e análise dos dados
- Técnicas de regressão avançadas
- Agrupamento
- Classificação
- Associação



Encerramento

- Materiais do curso: <https://github.com/gapacheco/ICD.git>
- Formulário de sugestões: <https://forms.gle/YdMY4TnVeNHSicR6>



Contato

- Email: gacampacheco@gmail.com
- LinkedIn: www.linkedin.com/in/gabriel-oc-pacheco