

# O Que é Big Data?

EST0133 - Introdução à Modelagem de Big Data

---

Marcus Nunes

<https://introbigdata.org/>

<https://marcusnunes.me/>

Universidade Federal do Rio Grande do Norte

O Que é Big Data?

# O Que é Big Data?

- Em geral, diz respeito a áreas do conhecimento onde as ferramentas de análise de dados tradicionais não são a melhor escolha possível
- Big Data são os dados que possuem 5 V:
  - Volume
  - Velocidade
  - Variedade
  - Veracidade
  - Valor

# O Que é Big Data?

- Uma outra definição de Big Data se vale da Estatística para ser formulada
- Podemos considerar um conjunto de dados como Big Data se o tempo que levamos para ajustar um modelo aos dados é maior do que o tempo utilizado para a escolha deste modelo

# O Que é Big Data?

- Mike Franklin, da Universidade de Berkeley, diz o seguinte:
- “Big Data é todo conjunto de dados caro para manter e manipular e de onde é difícil extrair informações”
- Esta definição é relativa: para alguns, dados na casa dos terabytes podem ser caros para manter; para outros, dados na casa dos petabytes podem ser baratos para manter

Que Problemas Podem ser Resolvidos com Big  
Data?

# Que Problemas Podem ser Resolvidos com Big Data?

- Considere três tamanhos diferentes de loja:
  - Loja Pequena (comércio de bairro)
  - Loja Média (Nordestão)
  - Loja Grande (Amazon)
- Qual é o tipo de relacionamento que cada uma destas lojas possui com seus clientes?

## Que Problemas Podem ser Resolvidos com Big Data?

- Um cliente recorrente de uma loja pequena recebe tratamento personalizado, diferente de um cliente eventual
- Um cliente recorrente de uma loja média recebe tratamento padrão, idêntico ao de um cliente eventual
- Um cliente recorrente de uma loja grande recebe tratamento personalizado, diferente de um cliente eventual



# Que Problemas Podem ser Resolvidos com Big Data?

- Qual seu mecanismo de busca favorito? Google ou Bing? Por quê?
- Mesmo que o algoritmo do Bing seja superior ao do Google (e não sabemos disso na realidade), o quantidade de dados do Google faz ele ter resultados melhores
- E quanto mais gente usar o Google, melhor ele se tornará

# Que Problemas Podem ser Resolvidos com Big Data?

- Queremos mais dados para gerar previsões melhores
- Como prever o resultado das próximas eleições?
- Definir tamanho da amostra, entrevistar eleitores, tabular os dados e fazer a análise estatística
- Agregar pesquisas já realizadas e combinar os seus resultados:
  - <https://fivethirtyeight.com/politics/elections/>
  - <http://pollingdata.com.br/>

# Que Problemas Podem ser Resolvidos com Big Data?

- Em geral, queremos resolver problemas de *machine learning* (ou aprendizagem de máquina)
- Queremos ensinar o computador a aprender com os dados
- O objetivo é encontrar padrões nos dados que permitam prever observações futuras

Quem Trabalha com Big Data?

# Quem Trabalha com Big Data?

- Competências de um profissional 100% capacitado para trabalhar com Big Data:
  - Estatística
  - Programação
  - Negócios
  - Conhecer bem a área de atuação (internet, marketing, área financeira etc)

# Quem Trabalha com Big Data?

- Que tipo de profissionais temos no momento?
  - Bons estatísticos e matemáticos que escrevem códigos sem otimização
  - Bons cientistas da computação que entendem um pouco de estatística e matemática
  - Bons cientistas da computação que entendem um pouco de negócios, depois de muita experiência na área
  - Especialistas em alguma área de atuação
  - Gerentes que sabem fazer estas pessoas trabalharem juntas

# Quem Trabalha com Big Data?

- Estatísticos
- Programadores
- Físicos
- Cientistas de Dados

# Quem Trabalha com Big Data?

- Cientista de Dados (*Data Scientist*) é o novo nome para Estatístico
- Alguns dizem que o Cientista de Dados é um Estatístico que mora em São Francisco e usa um Mac
- No fundo, ambos são a mesma coisa, embora uma destas profissões trabalhe melhor seu marketing pessoal



# O Que é Big Data?

EST0133 - Introdução à Modelagem de Big Data

---

Marcus Nunes

<https://introbigdata.org/>

<https://marcusnunes.me/>

Universidade Federal do Rio Grande do Norte