

Big Data, Ciência de Dados e Ciências Atuariais

X Simpósio de Atuária

Marcus Nunes

8 de Abril de 2021

Departamento de Estatística - UFRN

Quem sou Eu?

Quem sou Eu?

- Marcus Nunes, professor do Departamento de Estatística da UFRN
- Vice-Coordenador do Laboratório de Estatística Aplicada - <http://lea.estatistica.ccet.ufrn.br/>
- Curso de Modelagem de Big Data - <https://introbigdata.org/>
- Aplicações em Shiny - <http://shiny.estatistica.ccet.ufrn.br/>
- Site Pessoal - <https://marcusnunes.me/>

Sobre o que Conversaremos Hoje?

Sobre o que Conversaremos Hoje?

- Big Data, Ciência de Dados e Ciências Atuariais
- Ideias conceituais
- Foco maior em Ciência de Dados
- Uma hora de apresentação, meia hora de conversa

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?

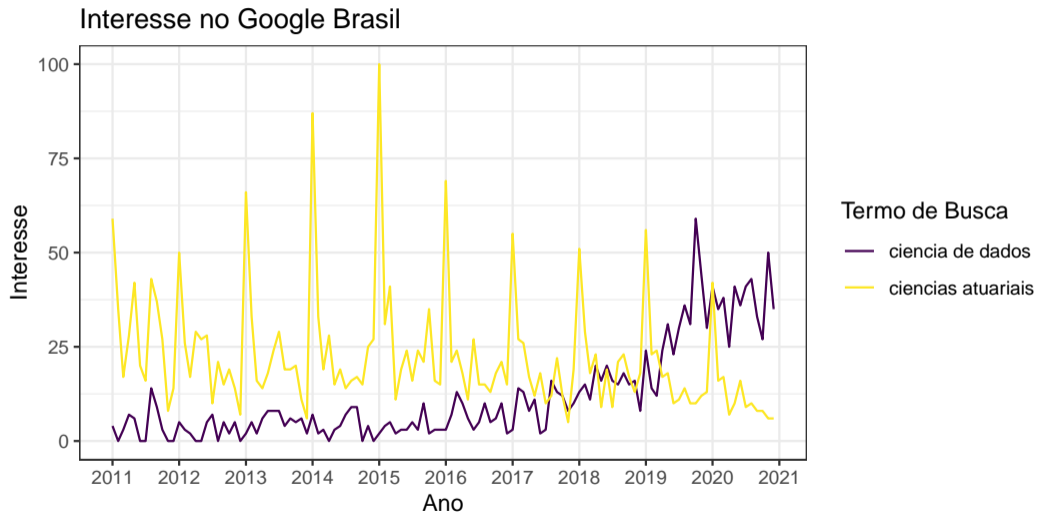
- Mike Franklin, da Universidade de Berkeley, diz o seguinte:
- “Big Data é todo conjunto de dados caro para manter e manipular e de onde é difícil extrair informações”
- Esta definição é relativa: para alguns, dados na casa dos terabytes podem ser caros para manter; para outros, dados na casa dos petabytes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

O que é Ciência de Dados?

O que é Ciência de Dados?

- *Buzzword* muito utilizada atualmente
- Juntamente com *big data* e *data science*, o termo tem ganhado muita força nos últimos anos

O que é Ciência de Dados?



O que é Ciência de Dados?



O que é Ciência de Dados?

- Alguém tem alguma definição?

O que é Ciência de Dados?

- Multidisciplinaridade: estatística, computação e comunicação
- Uso intensivo do computador
- Métodos de clusterização, classificação e regressão
- Predição, não inferência
- Extrair conhecimento dos dados

O que é um Cientista de Dados?

O que é um Cientista de Dados?



O que é um Cientista de Dados?

- Cientista de Dados (*Data Scientist*) é o novo nome para Estatístico
- Já foi dito 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 um Cientista de Dados?

- Estatísticos
- Programadores
- Físicos
- Cientistas de Dados

O que é um Cientista de Dados?

- É alguém que entende mais de programação do que um Estatístico tradicional
- É alguém que entende mais de estatística do que um Cientista da Computação tradicional
- E, principalmente, é alguém que consegue encontrar soluções para problemas juntando estas duas áreas do conhecimento

O que é um Cientista de Dados?

- Que área do conhecimento tem analisado e modelado dados a respeito de longevidade a fim de realizar previsões para o futuro?
- Dica: é a mesma área do conhecimento que faz modelagem de risco para seguros, crédito e fundos de pensão

O que é um Cientista de Dados?

- Que área do conhecimento tem analisado e modelado dados a respeito de longevidade a fim de realizar previsões para o futuro?
- Dica: é a mesma área do conhecimento que faz modelagem de risco para seguros, crédito e fundos de pensão
- Ciências atuariais

Quem já jogou RPG?

Quem já jogou RPG?

FORÇA

DISTREZA

CONSTITUIÇÃO

INTELIGÊNCIA

SABEDORIA

CARISMA

INSPIRAÇÃO

BÔNUS DE PROFICIÊNCIA

Força
 Destreza
 Constituição
 Inteligência
 Sabedoria
 Carisma

TESTES DE RESISTÊNCIA

Acrobacia (Des)
 Arcanismo (Int)
 Atletismo (For)
 Atuação (Car)
 Blefar (Car)
 Furtividade (Des)
 História (Int)
 Intimidação (Car)
 Intuição (Sab)
 Investigação (Int)
 Lidar com Animais (Sab)
 Medicina (Sab)
 Natureza (Int)
 Percepção (Sab)
 Persuasão (Car)
 Prestígio (Des)
 Religião (Int)
 Sobrevivência (Sab)

PERÍCIAS

CLASSE ARM.

INICIATIVA

DESLOC.

PV Totais

PONTOS DE VIDA ATUAIS

PONTOS DE VIDA TEMPORÁRIOS

Total

DADOS DE VIDA

SUCESSOS ○○○○

FRACASSOS ○○○○

TESTES CONTRA A MORTE

TRAÇOS DE PERSONALIDADE

IDEIAS

LIGAÇÕES

DEFEITOS

| NOME | BÔNUS | DANO / TIPO |
|------|-------|-------------|
| | | |
| | | |
| | | |

ATAQUES E MAGIAS



O que um Cientista de Dados Precisa Saber?

ESTatística: coletar, organizar, analisar e interpretar dados, provar teoremas e manipular expressões algébricas

EST

O que um Cientista de Dados Precisa Saber?

SIStemas: ser proficiente em computação em geral, sabendo como usar diferentes programas e sistemas operacionais

EST

SIS

O que um Cientista de Dados Precisa Saber?

ALGORITMOS: ser capaz de traduzir procedimentos e instruções para a realização de uma tarefa em alguma linguagem de programação

EST

SIS

ALG

O que um Cientista de Dados Precisa Saber?

COMunicação: entender problemas de outras áreas e comunicar suas conclusões para outras pessoas

EST

SIS

ALG

COM

O que um Cientista de Dados Precisa Saber?

PERsistência: tentar diferentes maneiras de encarar os problemas, mesmo quando eles parecem sem solução

EST

SIS

ALG

COM

PER

O que um Cientista de Dados Precisa Saber?

SORte: estar no lugar certo e na hora certa
e ter as habilidades necessárias
quando esta hora chegar

EST

SIS

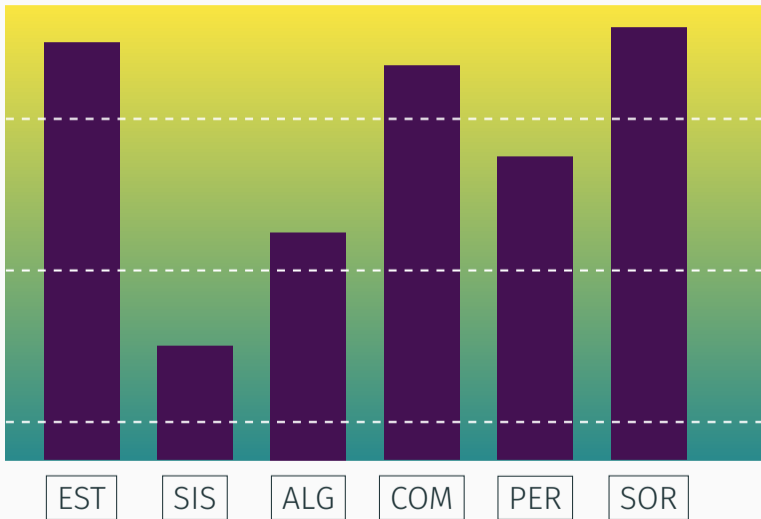
ALG

COM

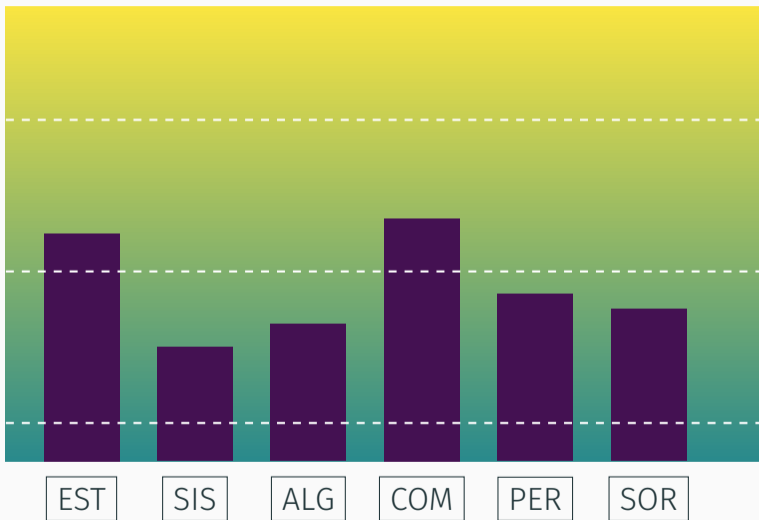
PER

SOR

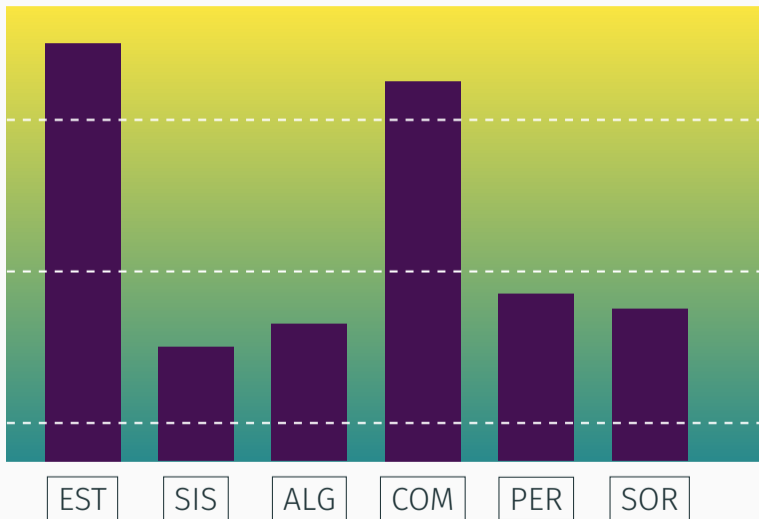
O que Eu Sei



Não é Bom Estar na Média



Seja Muito Bom em Algumas Áreas



Mas e as Ciências Atuariais?

Mas e as Ciências Atuariais?

- Maximizem suas qualidades e diferenciais
- Aproveitem as ferramentas de Ciência de Dados
- Reavaliem as técnicas já utilizadas na área e vejam como elas podem ser aprimoradas

Mas e as Ciências Atuariais?

- Seguro de carros
- Modelagem de mortalidade
- Sistemas de saúde
- Risco climático

Mas e as Ciências Atuariais?

- Há muito material disponível online
- Boa parte dele está gratuito
- A maioria é em inglês

Aplicações

Target

- How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



Fonte:

[http://www.forbes.com/sites/kashmirhill/2012/02/16/
how-target-figured-out-a-teen-girl-was-pregnant-before-her-](http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-)

- Compartilhamos dados com empresas sempre que fazemos compras
- As empresas usam estes dados para traçar o perfil dos clientes
- Com isto, podem enviar ofertas direcionadas para cada segmento de público

- Certa vez, um pai chegou furioso a uma loja da Target em Minneapolis, exigindo falar com o gerente
- Ele argumentava que a filha recebeu, pelo correio, ofertas de roupas de bebê e berços
- Detalhe: a filha ainda era aluna do Ensino Médio

- Na lógica do pai, isto estaria encorajando a filha a engravidar
- O gerente da Target se desculpou e disse que ia averiguar o que tinha acontecido
- Nada de estranho foi descoberto; mesmo assim, o gerente ligou para o pai uns dias depois e se desculpou novamente

Target

- Exemplo de perfil traçado pela Target:
- Imagine uma cliente chamada Jenny Ward, 23 anos, moradora de Atlanta
- Se em março ela comprou
 - hidratante de manteiga de cacau
 - uma bolsa grande
 - suplementos de zinco e magnésio
 - um tapete azul claro
- Segundo o modelo da loja, há 87% de chance de ela estar grávida e dar a luz no fim de agosto

- No caso específico da Target, o modelo deles é capaz de enviar correspondência para as clientes grávidas em momentos bastante específicos da gestação
- Além disso, serve para indicar como organizar fisicamente as lojas (por exemplo, colocar fraldas ao lado de cerveja)
- É possível descobrir que, quando casam, as pessoas comprar tipos diferentes de cereal

- Conclusão: depois da desculpa por telefone do gerente, o pai voltou a ligar para a loja
- Desta vez, ele se desculpou e disse que “havia eventos em sua casa de que ele não estava completamente a par”
- De fato, a menina estava grávida e deu a luz poucos meses depois do contato da loja

- How Netflix Reverse Engineered Hollywood
- To understand how people look for movies, the video service created 76,897 micro-genres
- We took the genre descriptions, broke them down to their key words... and built our own new-genre generator

Fonte:

<http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>

- Gêneros de filmes em locadoras? Ação, Aventura, Comédia, Drama etc.
- Gêneros de filmes na Netflix? African-American Crime Documentaries, Scary Cult Movies from the 1980s, Feel-good Romantic Spanish-Language TV Shows, Visually Striking Latin American Comedies e mais
- Assim, é mais fácil sugerir filmes que podem agradar o usuário

- Serviço de streaming de música
- 144 milhões de assinantes, 320 milhões de usuários ativos (Outubro de 2020)



- Terabytes de dados são produzidos pelos usuários
- É possível relacionar dados dos usuários para sugerir novos artistas e canções
- Críticos não são necessários neste sistema: novas sugestões são criadas a partir de algoritmos

Fonte: <https://labs.spotify.com/tag/big-data/>

E a UFRN?

E a UFRN?

- Berkeley possui uma disciplina chamada Data 8: Foundations of Data Science - <http://data8.org/>
- Três perspectivas: pensamento inferencial, pensamento computacional e relevância no mundo real
- Data science minor approved by multiple UC Berkeley colleges - <https://www.dailycal.org/2019/10/07/data-science-minor-approved-by-multiple-uc-berkeley-col>

E a UFRN?

- Ciência de Dados deveria ser uma iniciativa global da universidade
- A UFRN deveria ter uma equipe de profissionais trabalhando com os dados da universidade
- Além disso, o assunto Ciência de Dados deveria ter ao menos uma disciplina disponível para todas as graduações da universidade

Considerações Finais

Considerações Finais

- Dá pra fazer Ciência de Dados sozinho? Sim.
- É melhor fazer em equipe? Evidentemente que sim.
- Diferentes habilidades trarão conhecimentos diferentes para quem for trabalhar na área
- Tenham um portfólio online!

Big Data, Ciência de Dados e Ciências Atuariais

X Simpósio de Atuária

Marcus Nunes

8 de Abril de 2021

Departamento de Estatística - UFRN