Marcelo d'Almeida
md@id.uff.br

Vítor Lourenço
vitornaslourenco@gmail.com

CCD-UFF

Clube de Ciência de Dados

Universidade Federal Fluminense

CCD-UFF
Clube de Ciência de Dados

Roteiro

- Os porquês
- Motivação
- Quem participará
- Trajetória e estudos
- Tema proposto
- Propostas de projetos
- Fontes de dados
- Citações

Por que montar um “Clube”?

- Clube – “Associação de pessoas que têm por objetivo a consecução de determinado propósito ou fim comum.” [1]
- Explorar a integração entre alunos ao redor de um assunto em comum
- Possibilidade de estudar conteúdos não abordados pelo curso ou de aprofundar aqueles que já são
- Servir de inspiração à outras iniciativas de mesmo tipo

Por que “Ciência de Dados”?

- Ciência de Dados é um conjunto de conhecimentos – obtenção, organização, propriedade, processamento, análise e visualização – acerca de um conjunto de dados, utilizado para analisar sua interferência no meio em que foi gerado [2]
- Objetivo principal: projetar soluções de problemas através dos processos de obtenção, processamento, análise e visualização de conjuntos de dados

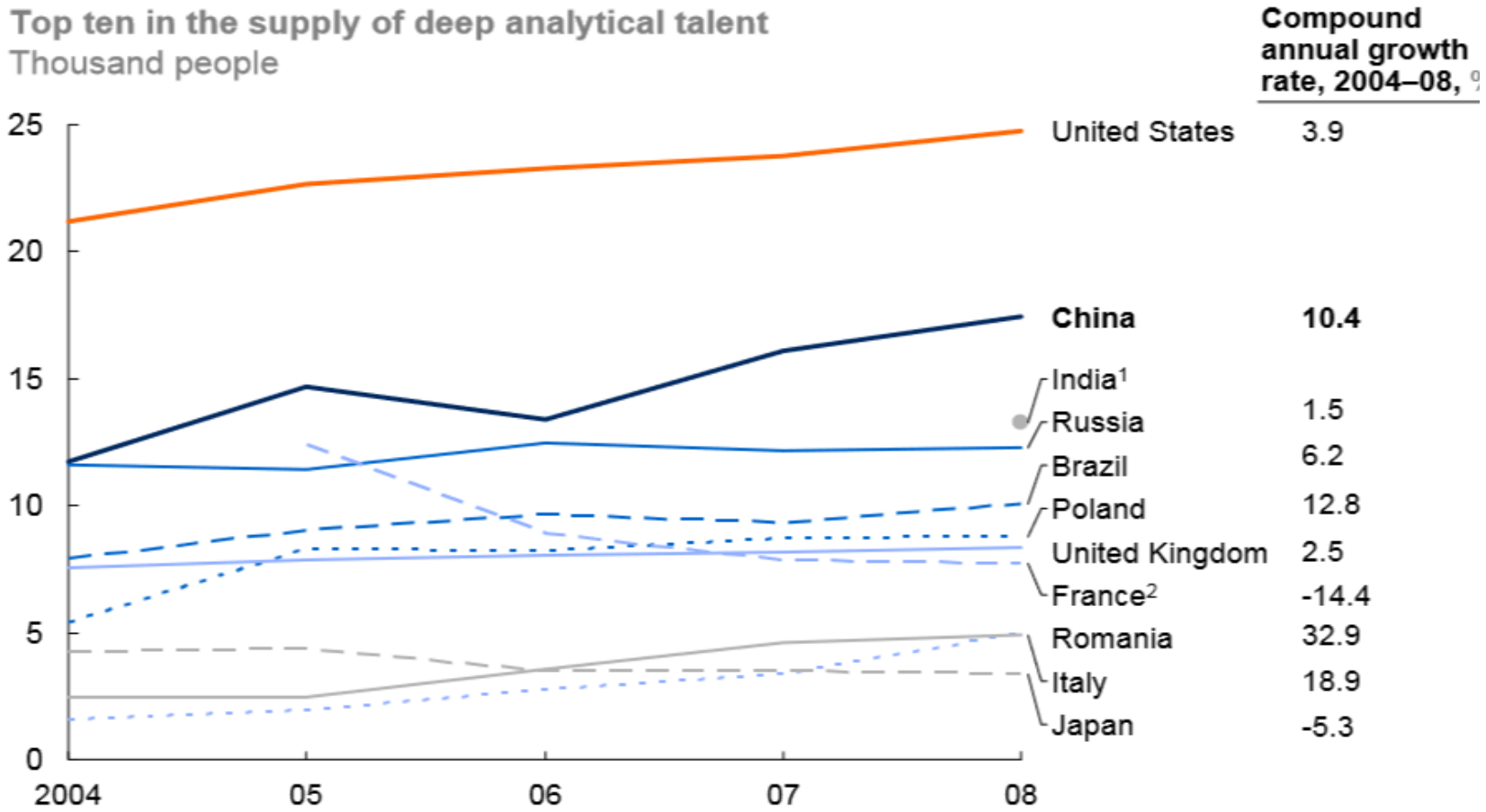
Por que “Ciência de Dados"?

- Grande demanda de *data scientists*
  - Crescimento do volume de dados globalmente
  - É estimado que, em 2018, o EUA, sozinho, enfrente uma escassez de 140,000 a 190,000 pessoas com *deep analytics skills* assim como 1.5 mi gerentes e analistas com conhecimentos em análise de *big data* para decisões mais efetivas [3]
- E grande salários

Por que “Ciência de Dados”?

O rápido crescimento dos dados globais [4]

Por que "Ciência de Dados"?



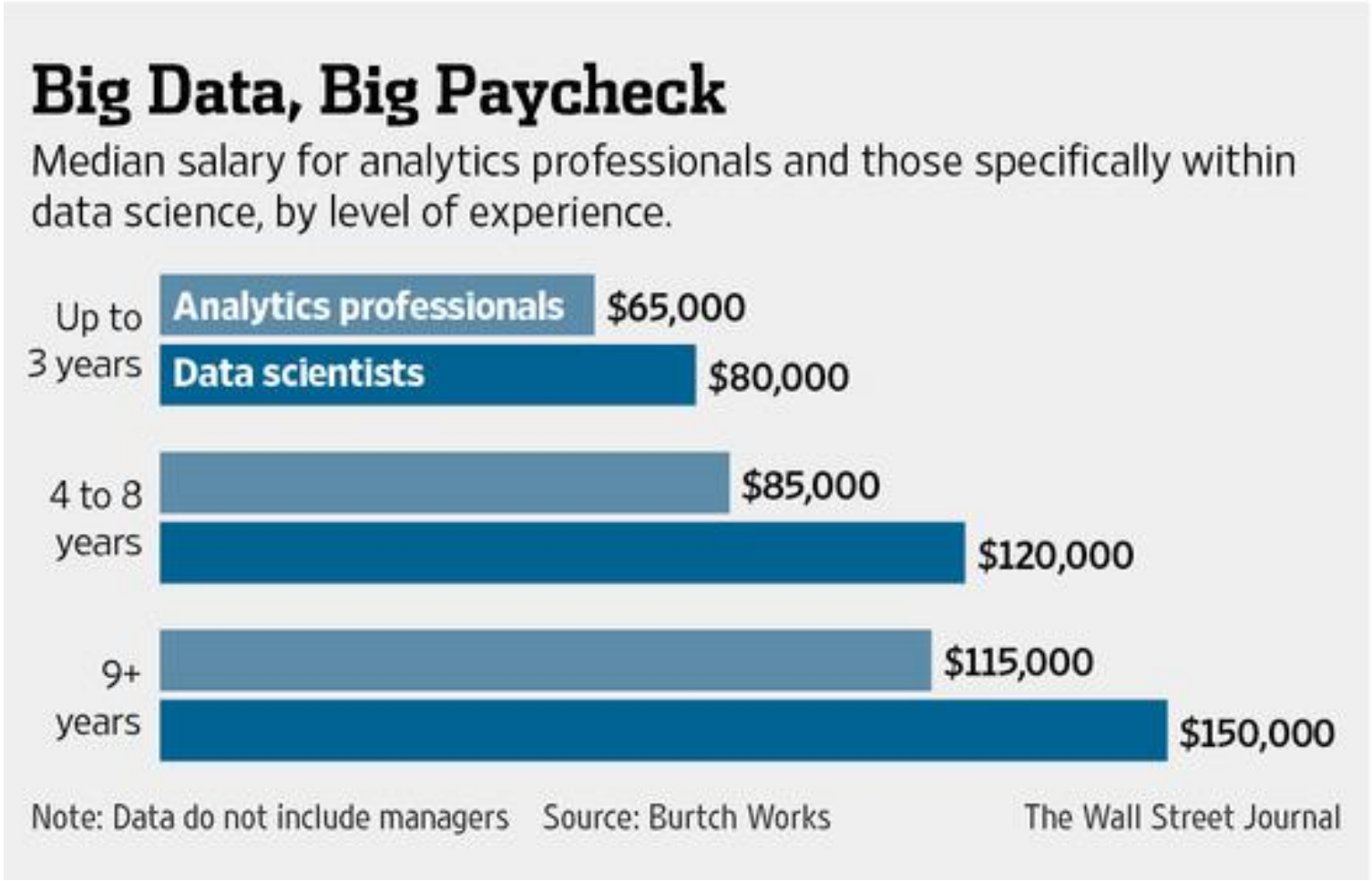
1 India ranked third in 2008 with 13,270 people with deep analytical skills but India does not have a time series for these data.
2 For France, the compound annual growth rate is for 2005 and 2008 because of data availability.
SOURCE: Eurostat; national statistical agencies; Japan Ministry of Education; India Sat; NASSCOM; China Statistical Yearbook; China *Education News*, April 2005; IMF *World Economic Outlook*; McKinsey Global Institute analysis

Top dez na oferta de *deep analytical talent* [3]

Por que “Ciência de Dados"?

- Grande demanda de *data scientists*
  - Crescimento do volume de dados globalmente
  - É estimado que, em 2018, o EUA, sozinho, enfrente uma escassez de 140,000 a 190,000 pessoas com *deep analytics skills* assim como 1.5 mi gerentes e analistas com conhecimentos em analise de *big data* para decisões mais efetivas [3]
- E grande salários

Por que "Ciência de Dados"?



Big Data, grande contracheque [5]

Qual é a motivação?

- Buscar por novos desafios
- Trabalhar em equipe
- Estimular a aprendizagem dinâmica no campo de Ciência de Dados
- Possibilidade de criação de artigos científicos e aplicações que beneficiem a comunidade como um todo

Qual é a motivação?

- Processo:

Que tipos de alunos participarão?

- Alunos de graduação, interessados em Ciência de Dados e aptos a propor e desenvolver soluções e aplicações para desafios reais
- Alunos de pós graduação interessados em auxiliar, por meio de consultoria ou tutoria, os projetos e seus desenvolvimentos
  - Banco de Dados
  - Computação Gráfica
  - Engenharia de Software
  - Inteligência Artificial

Trajetória e estudos

- Coursera.org
  - Johns Hopkins University
    - Data Science Specialization
  - University of Illinois at Urbana-Champaign
    - Data Mining Specialization
  - University of Washington
    - Introduction to Data Science
- Livros
  - Data Mining Concepts and Techniques
  - Data Mining Practical Machine Learning Tools and Techniques

Data Science Specialization

1. The Data Scientist's Toolbox
2. R Programming
3. Getting and Cleaning Data
4. Exploratory Data Analysis
5. Reproducible Research
6. Statistical Inference
7. Regression Models
8. Practical Machine Learning
9. Developing Data Products

Instructors

Brian Caffo
Professor
Department of Biostatistics

Jeff Leek
Assistant Professor
Biostatistics

Roger D. Peng
Associate Professor
Biostatistics

Data Science Specialization – Coursera – Johns Hopkins University [6]

Data Mining Specialization

Instructors

1. Pattern Discovery in Data Mining
2. Text Retrieval and Search Engines
3. Cluster Analysis in Data Mining
4. Text Mining and Analytics
5. Data Visualization

Jiawei Han
Abel Bliss Professor
Department of Computer Science

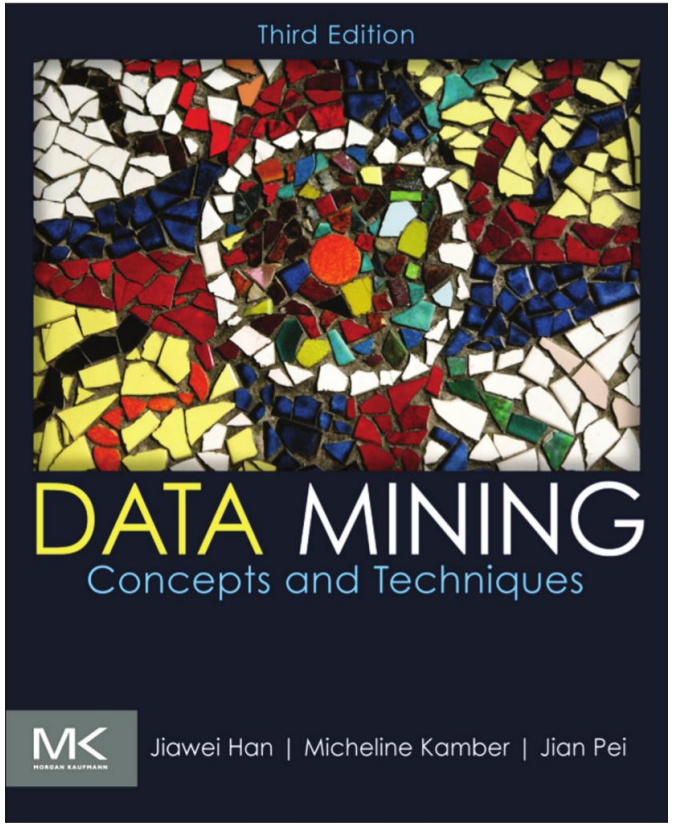
ChengXiang Zhai
Professor
Department of Computer Science

John C. Hart
Professor
Department of Computer Science

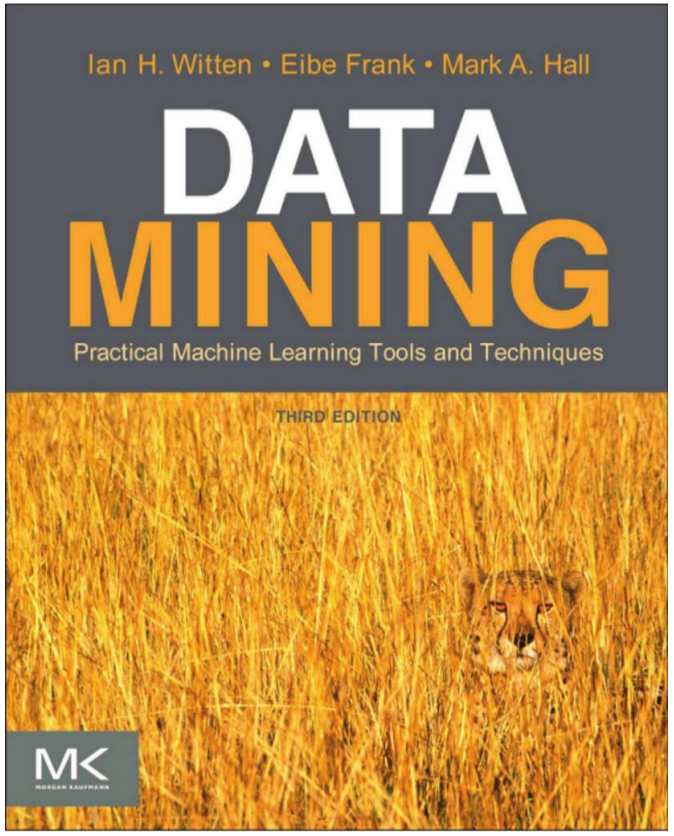
ILLINOIS

Data Mining Specialization – Coursera – University of Illinois at Urbana-Champaign [7]

Livros



Data Mining: Concepts and Techniques [8]



Data Mining: Practical Machine Learning Tools and Techniques [9]

Tema proposto: *Smart City*

- Cidades resilientes
  - Adaptabilidade e flexibilidade às circunstancias hostis
- *Smart Cities*
  - Economia, mobilidade, ambiente e governança inteligentes
  - Conectar pessoas à pessoas e à serviços em comunicações inteligentes

Smart City [10]

Propostas de projetos

- *Smart Cities*
  - Meio ambiente
    - Análise de chuvas e níveis de rios para detecção de enchentes
    - Análise dos níveis de poluição da água e do ar
  - Educação
    - Análise de desempenho escolar em relação a taxa aluno/professor
  - Saúde
    - Análise meteorológica e casos de dengue
    - Análise das condições de saúde das mães e sua relação com a saúde do feto/bebê
  - Transporte e Mobilidade
    - Análise das ocorrências de trânsito e suas causas
  - Segurança pública
    - Análise das ocorrências e suas reincidências ao longo de períodos do ano

Fontes de dados

- Rio Datamine

Rio Datamine [11]

- Armazém de Dados

Armazém de Dados [12]

- Data.Rio

Data.rio [13]

Citações

- Houaiss, A. Dicionário Houaiss da Língua Portuguesa. 1 ed. Rio de Janeiro: Rio de Janeiro, 2001. [1]
- V. Dhar, "Data science and prediction", Commun. ACM, vol. 56, no 12, p. 64–73, dez. 2013. [2]
- McKinsey Global Institute, MGI Big Data full report [3]
- http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode [4]
- http://www.wsj.com/articles/SB10001424052702304819004579489541746990638 [5]
- https://www.coursera.org/specialization/jhudatascience/1?utm_medium=listingPage [6]
- https://www.coursera.org/specialization/datamining/20?utm_medium=listingPage [7]

Citações

- Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber – Morgan Kaufmann Publishers, 3rd ed., 2011 [8]
- Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten, Eibe Frank – Morgan Kaufmann Publishers, 3rd ed., 2011 [9]
- http://smartcities.ieee.org/about.html [10]
- http://riodatamine.com.br/#/homepage [11]
- http://www.armazemdedados.rio.rj.gov.br/ [12]
- http://data.rio.rj.gov.br/ [13]

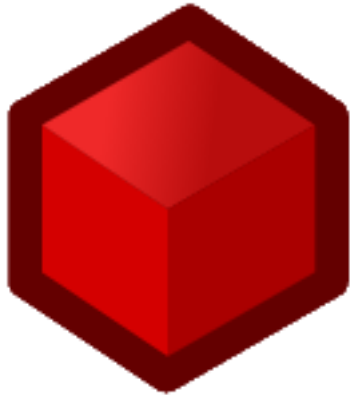
Obrigado!

Dúvidas?

Sugestões?

Marcelo d'Almeida
md@id.uff.br

Vítor Lourenço
vitornaslourenco@gmail.com



CCD-UFF
Clube de Ciência de Dados