

*Bicentenário da Independência:*  
200 anos de Ciência, Tecnologia e Inovação no Brasil

**DAL**  
&  
*Comunidade*

## CIÊNCIA DE DADOS NOS EIXOS CIÊNCIAS, EMPRESAS E GOVERNOS



**CEFET/RJ**

Eduardo Ogasawara  
eogasawara@ieee.org  
<https://eic.cefet-rj.br/~eogasawara>

# Ciência de Dados

- Ciência de dados é a Ciência que estuda a análise de dados
  - A extração de conhecimento a partir de dados (Big Data)
- Envolve
  - Análise de dados
  - Aprendizagem estatística e de máquina
  - Gerência de grandes coleções de dados
  - Mineração de dados
  - Computação paralela e distribuída
  - Visualização de dados
- Característica multidisciplinar
- Aplicações na ciência/indústria/governo
  - Bioinformática, petróleo, energia, finanças, astronomia, Internet, mobilidade urbana, defesa cibernética, educação, etc



[1] V. Dhar, 2013, Data science and prediction, *Communications of the ACM*, v. 56, n. 12, p. 64–73.

[2] L. Cao, 2017, Data science: A comprehensive overview, *ACM Computing Surveys*, v. 50, n. 3

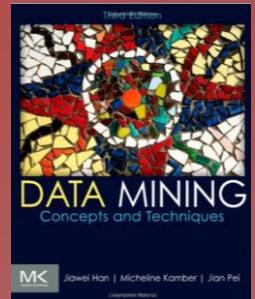
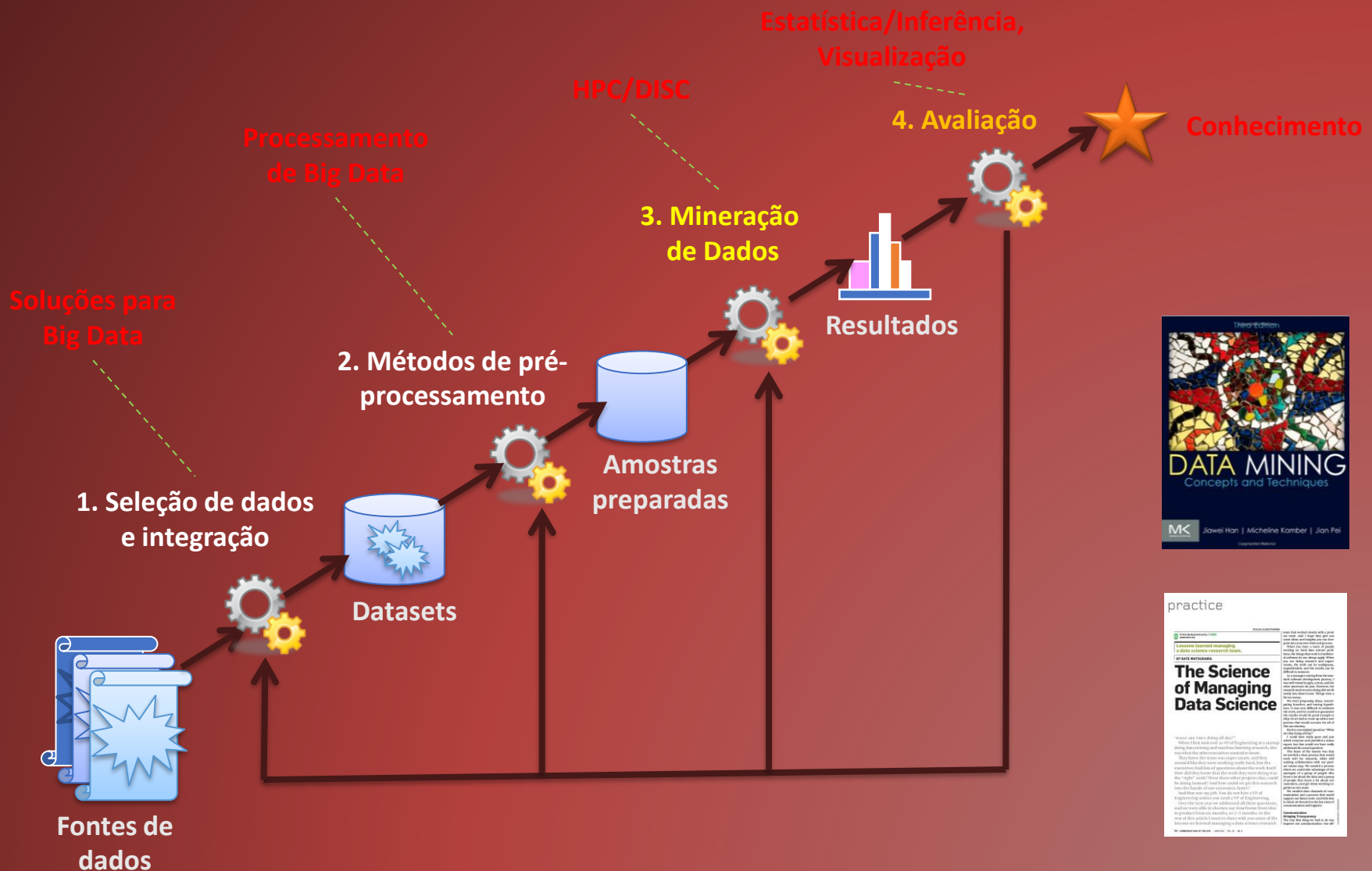
[3] L. Cao, 2017, Data science: Challenges and directions, *Communications of the ACM*, v. 60, n. 8, p. 59–68.

# Por que Ciência de Dados?

- Cenário de Big Data:
  - O crescimento explosivo dos dados: de terabytes a petabytes
  - Coleta de dados e disponibilidade de dados
- Principais fontes de dados abundantes e diversos
  - Negócios: comércio eletrônico, transações
  - Ciência: sensores, astronomia, bioinformática, simulação
  - Sociedade e todos: dados abertos, IoT, sensoriamento populacional
- Estamos nos afogando em dados, mas famintos por conhecimento!
  - Mineração de Dados: Análise automatizada de grandes conjuntos de dados
  - Ciência de Dados: traz a Ciência para a Análise dos Dados



# Processo de Descoberta de Conhecimento

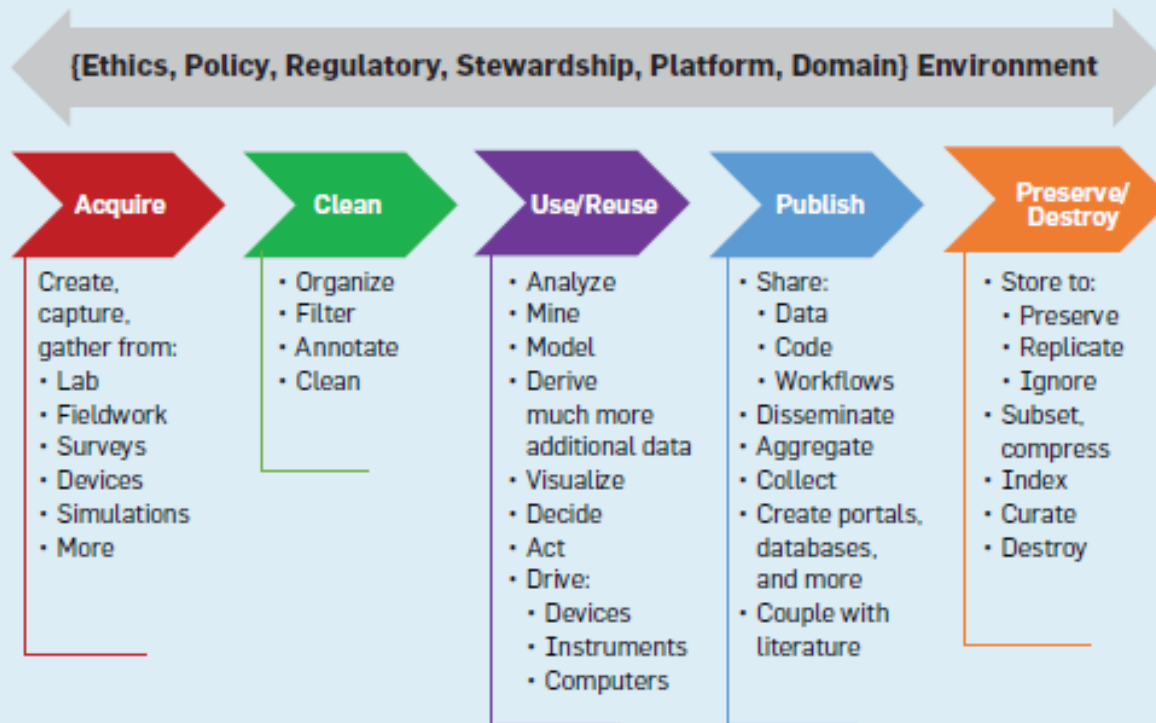


[1] K. Matsudaira, 2015, The science of managing data science, *Communications of the ACM*, v. 58, n. 6, p. 44–47.

[2] J. Han, J. Pei, and M. Kamber, 2011, *Data Mining: Concepts and Techniques*. Elsevier.

# Ciclo de Vida da Ciência de Dados

Figure 1. Example of a data life cycle and surrounding data ecosystem (reprinted with permission).<sup>1</sup>



contributed articles

**A cycle that traces ways to define the landscape of data science.**

BY VICTORIA STODDEN

## The Data Science Life Cycle: A Disciplined Approach to Advancing Data Science as a Science

THE SCIENCE AND RESEARCH ENTERPRISE IS INVESTING opportunities to accelerate science and discovery afforded by computational and data-enabled technologies, often broadly referred to as data science. For years ago, we see that an “academic” (of a scientific researcher) depicts a computer jobber working at all hours in search of a computer screen. “Since then, the use of data and computation has exploded in academic and industry research, and interest in data science is widespread in universities and institutions. Two key questions emerge for the research enterprise: How to train

**Key Insights**

- The data science life cycle is a disciplined approach to advancing data science as a science.
- The data science life cycle is a disciplined approach to advancing data science as a science.
- The data science life cycle is a disciplined approach to advancing data science as a science.

[1] V. Stodden, 2020, The data science life cycle, *Communications of the ACM*, v. 63, n. 7, p. 58–66.

[2] K. Matsudaira, 2015, The science of managing data science, *Communications of the ACM*, v. 58, n. 6, p. 44–47.



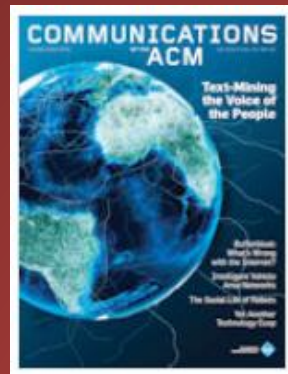
# Contexto de Ciência de Dados



ago/2002



dez/2008



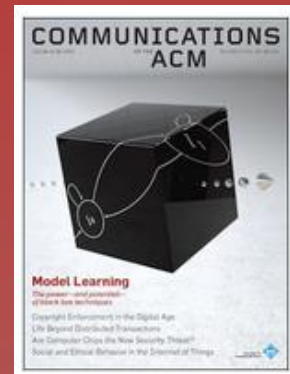
fev/2012



dez/2013



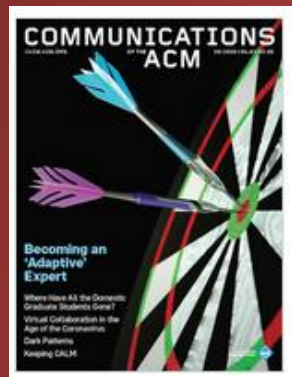
jul/2015



fev/2017



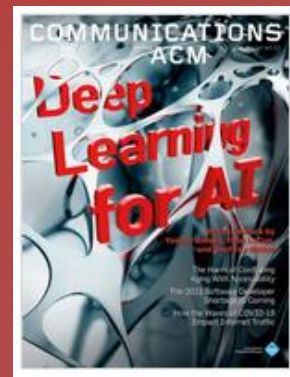
jul/2018



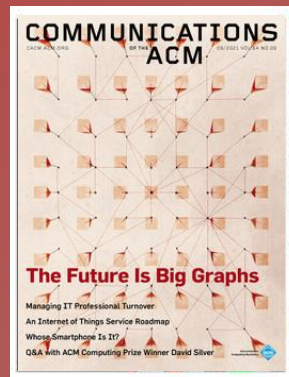
set/2020



mar/2021



jul/2021

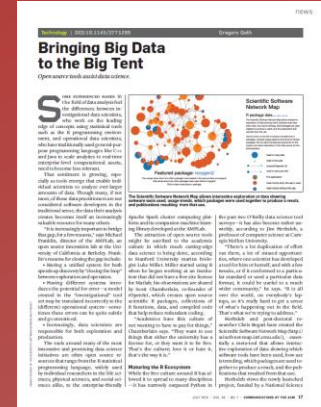
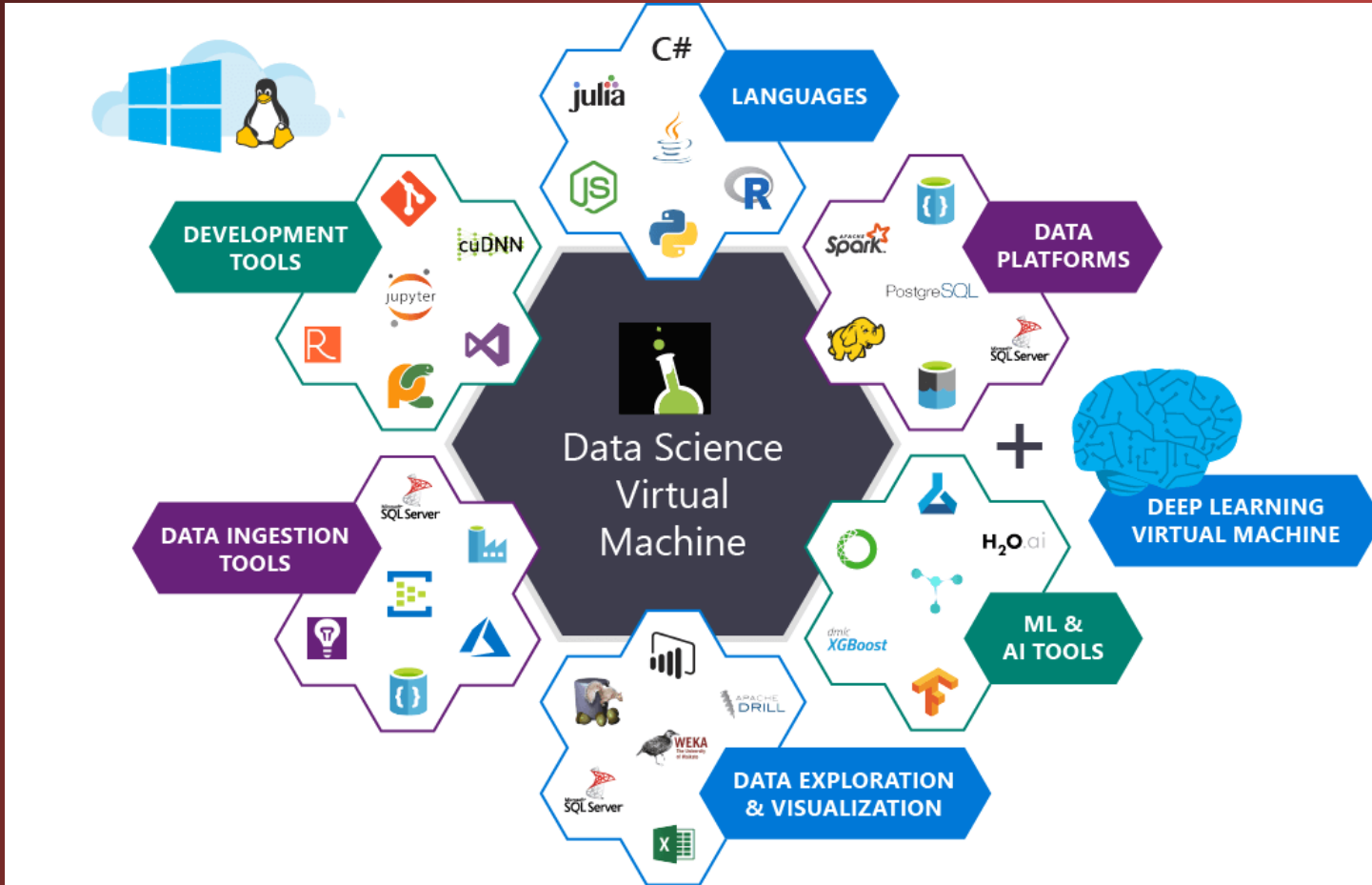


sep/2021



mar/2022

# Ecosystema de Ciência de Dados



[1] G. Goth, 2015, Bringing big data to the big tent, *Communications of the ACM*, v. 58, n. 7, p. 17–19.

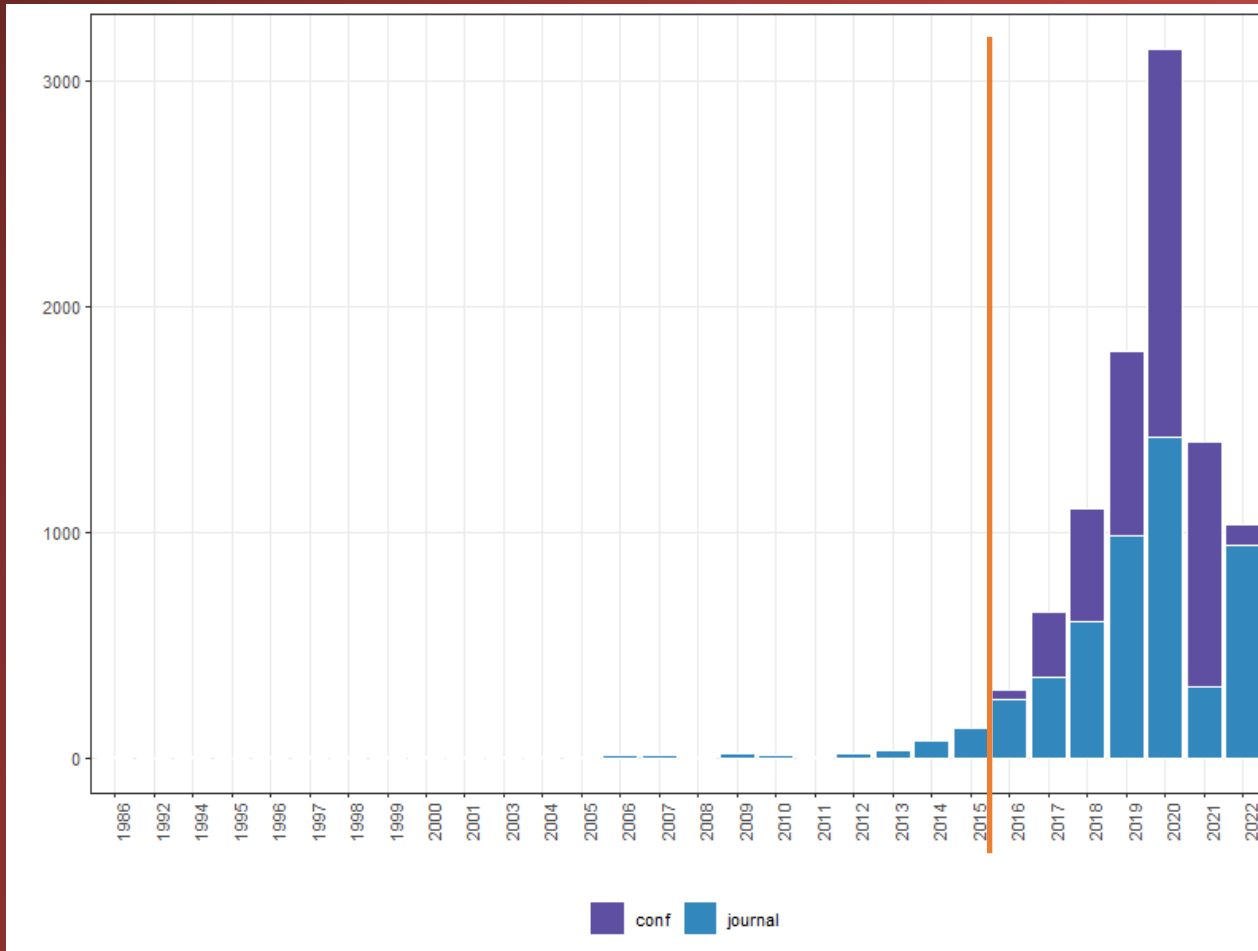
[2] <https://www.datasciencecentral.com/>

## *Perspectiva das Ciências*



## Como está área?

- Pesquisa básica e aplicada
- 9788 artigos publicados entre 1986 e 2022



# Como são as publicações de Ciência de Dados?

- Artigos de Review
- 926 artigos

1. Medicine	284	14. Energy	33
2. Computer Science	185	15. Business, Management and Accounting	31
3. Biochemistry, Genetics and Molecular Biology	171	16. Neuroscience	31
4. Mathematics	130	17. Pharmacology, Toxicology and Pharmaceutics	30
5. Engineering	115	18. Immunology and Microbiology	23
6. Decision Sciences	92	19. Earth and Planetary Sciences	22
7. Materials Science	81	20. Psychology	16
8. Environmental Science	63	21. Multidisciplinary	15
9. Chemistry	61	22. Arts and Humanities	11
10. Social Sciences	61	23. Nursing	11
11. Agricultural and Biological Sciences	57	24. Economics, Econometrics and Finance	8
12. Chemical Engineering	54	25. Veterinary	4
13. Physics and Astronomy	46	26. Dentistry	1



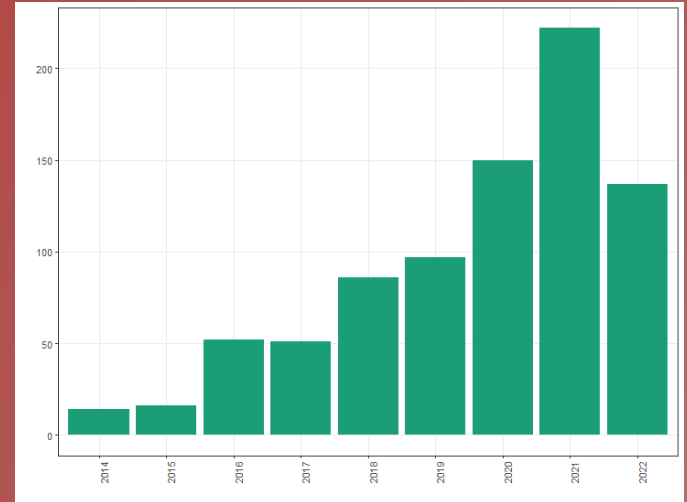
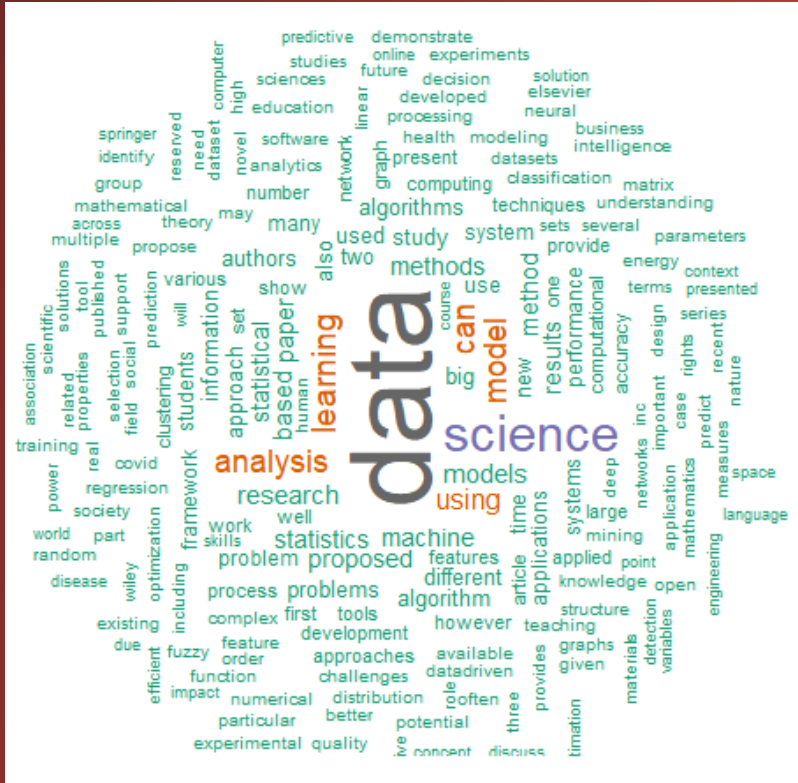






# Ciência de Dados na Matemática & Estatística

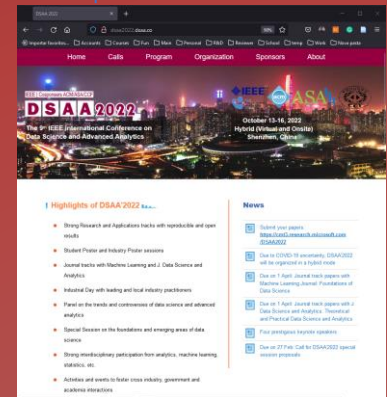
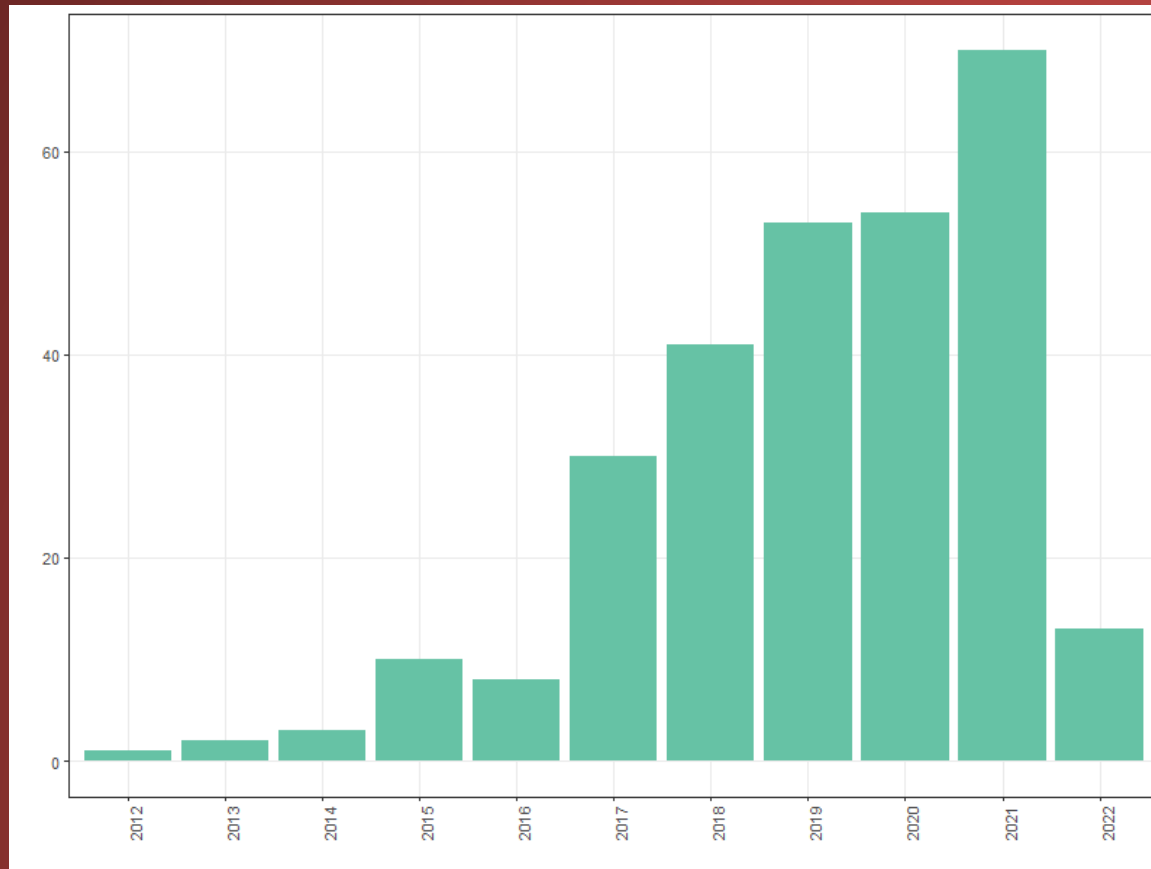
- Termos Estatísticos e Matemáticos
  - Modelos, Análise e Ciência
- Ciência para a Análise dos Dados (John Tukey – 50 anos)



[1] D. Donoho, 2017, 50 Years of Data Science, Journal of Computational and Graphical Statistics, v. 26, n. 4, p. 745–766. (cited 236)

# Existem eventos?

- 285 eventos entre 1986 e 2022
  - IEEE International Conference on Data Science and Advanced Analytics



*Quais são os desafios?*

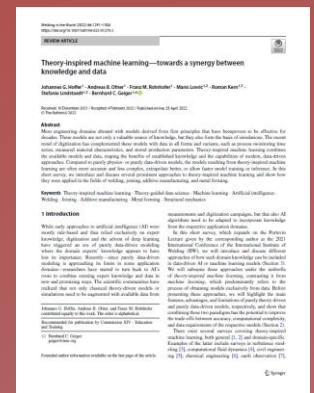
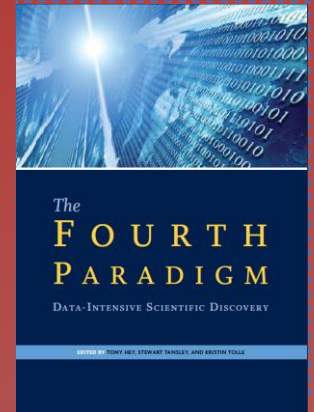
# Automatização

- A automatização em ciência de dados visa facilitar e transformar o trabalho dos cientistas de dados
- Mecanização
  - Ocorre quando uma tarefa é tão bem especificada que não há necessidade de envolvimento humano
  - AutoML / AutoWeka
- Composição
  - lida com o sequenciamento estratégico de tarefas ou integração de diferentes partes de uma tarefa
  - Workflows
- Assistência
  - deriva da produção de elementos como visualizações, padrões, explicações, entre outros, que são especificamente direcionados para apoiar a eficiência humana
- Na engenharia de dados há várias oportunidades
  - Organização, Qualidade e Transformação de Dados



# Modelos Data-Driven versus Theory-Driven

- Ciência da Computação: terceiro pilar da ciência
  - Teoria
  - Experimentação
  - Simulação (in-silico)
- Descoberta Científica baseadas em Dados
- Modelos orientados a dados
- *Data-driven e theory-driven*
  - Concorrência
  - Realimentação
  - Associação
    - Construção de modelos centrado em dados
    - Que respeitem “leis científicas”
    - *Physical informed Neural Networks*



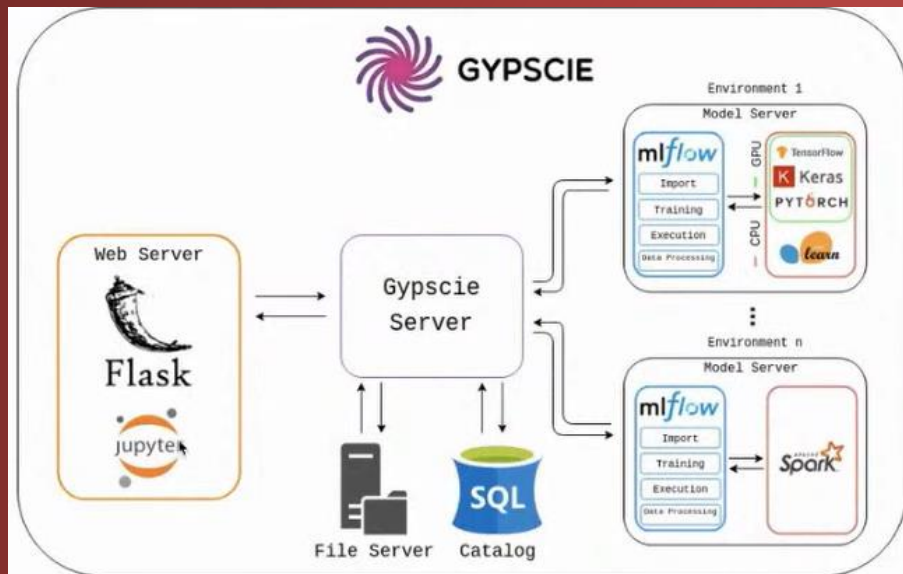
[1] T. Hey, 2009, *The Fourth Paradigm: Data-intensive Scientific Discovery*. Microsoft Research.

[2] J.G. Hoffer, A.B. Ofner, F.M. Rohrhofer, M. Lovrić, R. Kern, S. Lindstaedt, and B.C. Geiger, 2022, Theory-inspired machine learning - towards a synergy between knowledge and data, *Welding in the World*, v. 66, n. 7, p. 1291-1304.



# Gerência de Modelos

- Gerência de Modelos
  - Construção
  - Produção
  - Reutilização
    - Transferência de conhecimento
  - Seleção dinâmica
    - Distribuição
    - Espaço-tempo



**Ciclo de Seminários Online** Instituto de Computação

**Gypscie: Um Ambiente para Gerência de Dados e Modelos de Aprendizado de Máquina**

**Fabio Porto**  
LNCC

**TRANSMISSÃO AO VIVO**

**15 de Outubro**  
17:00

**DEXL**  
DATA EXTREME LAB  
Laboratório Nacional de Computação Científica

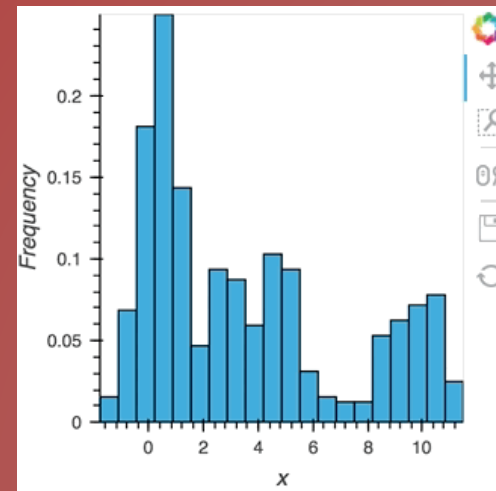
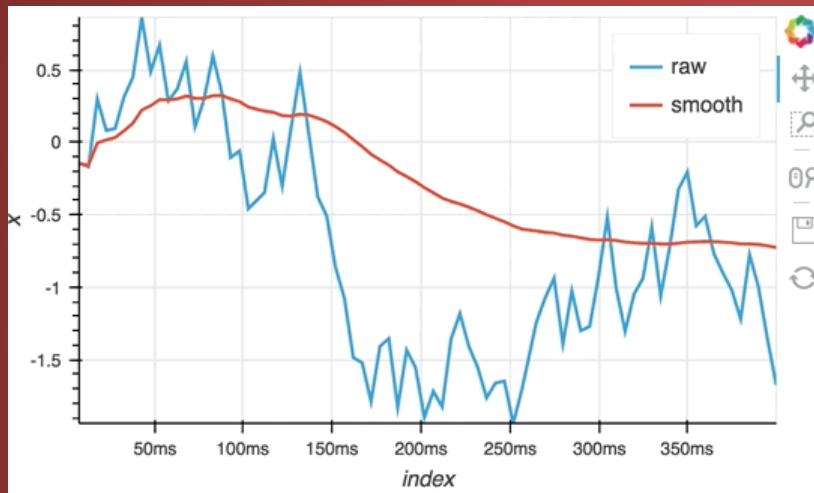
YouTube [http://tiny.cc/ucff\\_youtube](http://tiny.cc/ucff_youtube) <https://fb.me/informeic>

#InstitutoDeComputação #UFFContraCoronaVirus #FiqueEmCasa

[1] S. Idowu, D. Strüber, and T. Berger, 2022, Asset Management in Machine Learning: State-of-research and State-of-practice, ACM Computing Surveys (Jun.) <https://youtu.be/H2t3nPlnMl4>

# As múltiplas faces da não-estacionariedade

- Mudança de conceito (*Concept Drift*)
  - visão de Mineração de Dados
- Dilema da Plasticidade e Estabilidade
  - visão de Aprendizado de Máquina
- Padrões emergentes (*Emerging patterns*)
  - visão de Banco de Dados
- Crítica de Lucas
  - visão de Econometria



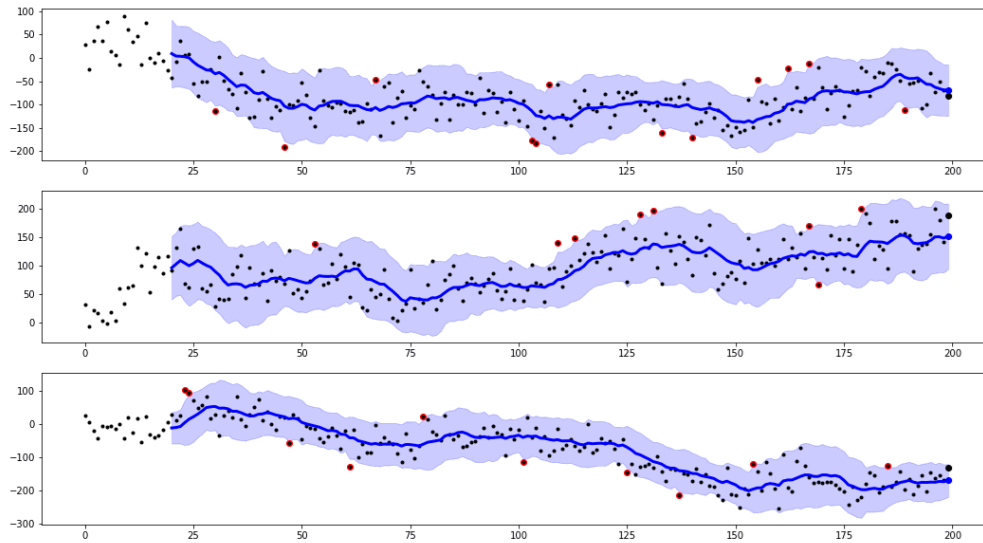
[1] A.S. Iwashita and J.P. Papa, 2019, An Overview on Concept Drift Learning, *IEEE Access*, v. 7, p. 1532–1547.

[2] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.

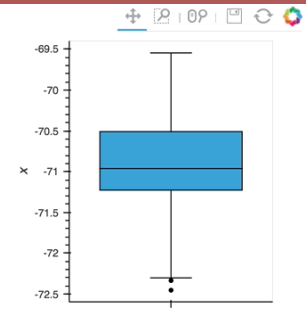
[3] D. Gujarati, 2002, *Basic Econometrics*. 4 ed. Boston; Montreal, McGraw-Hill/Irwin. / [4] S.O. Haykin, 2008, *Neural Networks and Learning Machines*. 3 ed. New York, Prentice Hall.

# Processamento de stream

- Processamento de stream
  - Velocidade
  - Volume de dados
- Detecção de eventos
  - Offline versus Online

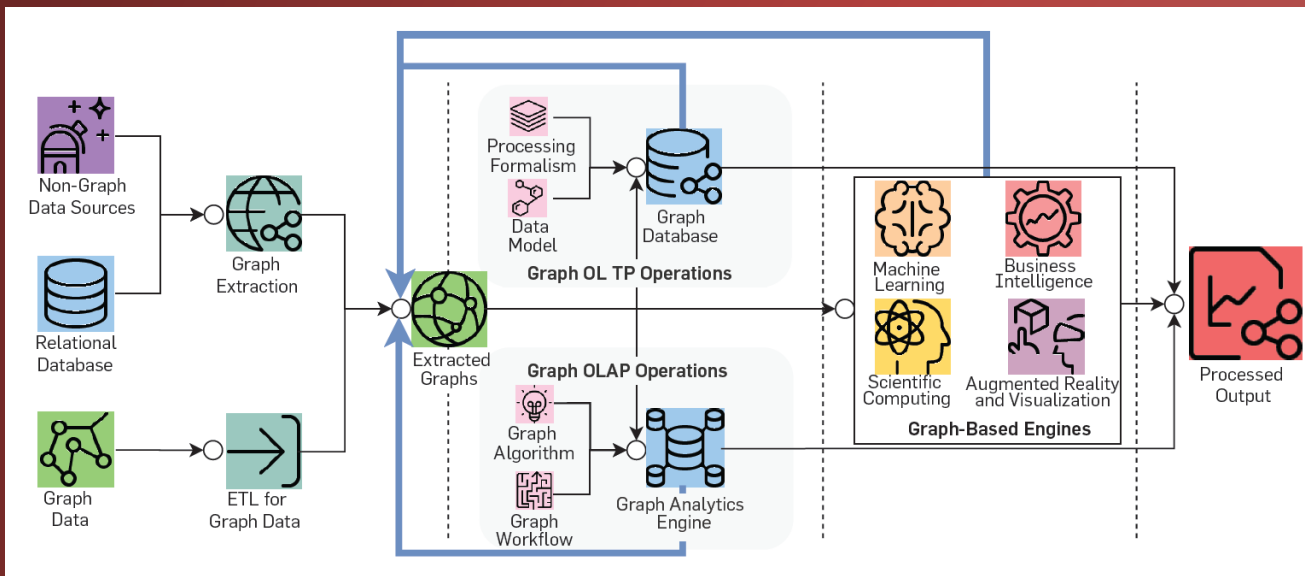
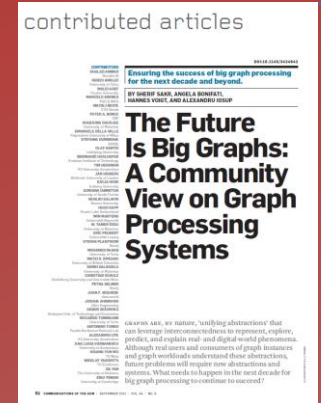


#	index	x
0	2017-10-27 19:43:04	-70.87262710547351
1	2017-10-27 19:43:04	-71.00788295730518
2	2017-10-27 19:43:04	-71.01297392873352
3	2017-10-27 19:43:04	-71.4148637783796
4	2017-10-27 19:43:04	-71.6089069520968
5	2017-10-27 19:43:04	-71.32610485802545
6	2017-10-27 19:43:04	-71.1935680343768
7	2017-10-27 19:43:04	-70.76579342173753
8	2017-10-27 19:43:04	-70.59743524950701
9	2017-10-27 19:43:04	-70.99300214112863



# Big Graph

- Grande número (mais de 60.000) de downloads em um ano do livro do Neo4j Graph Algorithms
- Grande interesse no uso de técnicas de grafos nas áreas de inteligência artificial e aprendizado de máquina
- Iniciativas de análise de propagação da covid-19 usando grafos



Ciclo de Seminários Online

Instituto de Computação

Projeto de Bancos de Dados NoSQL

Prof. Ronaldo Mello

05/04 - 16:00

UNIVERSIDADE FEDERAL DE SANTA CATARINA

INSTITUTO DE COMPUTAÇÃO

AVISANDO AO SEU DE

YouTube

f

Twitter

*Empresas*



# Características e áreas

- Características das empresas
  - Contexto de Big Data
  - Business Intelligence & Data Analytics
  - Necessidade de Deployment

- Áreas

- Financeira (análise de risco)
- Saúde (melhoria de diagnósticos)
- Turismo (perfis, segmentação)
- Energia (reduzir risco, otimizar investimentos)
- Manufatura (Industria 4.0, logística, predição)
- Jogos (perfil do consumidor)
- Farmacêutica (planejamento clínico)

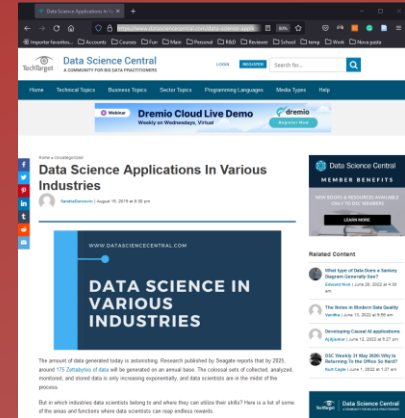


Figure 8: What Is the Main Driver for Your Data and/or Analytics Program?



Source: State of Data in Travel Survey, 2017

# Ciência de Dados nas empresas e indústrias

- Construir modelos úteis
- Desafios das soluções caixa preta
  - algoritmos, parâmetros de ajuste, efeitos sutis e montagens não são acessíveis
  - podem ser adequadas para soluções de curto prazo, mas a robustez à mudança não é certa
  - habilidades para compreender os métodos analíticos não são valorizadas e desenvolvidas



# Ciência de Dados: Carreira mais "sedutora" do século!

2012

Data scientists today are akin to the Wall Street "quants" of the 1980s and 1990s

In 2012, Harvard Business Review named data scientist the "sexiest job of the 21st century." More recently, Glassdoor named it the "best job of the year" for 2016.

"It isn't a big surprise," Dr. Andrew Chamberlain, Glassdoor's chief economist, told Business Insider. "It's one of the hottest and fastest growing jobs we're seeing right now."

According to Glassdoor, data scientists earn a base pay of **\$116,840 a year**, on average.



VFS Digital Design/Flickr

**Facebook:** \$133,841

**Apple:** \$149,963

**Airbnb:** \$117,229

**Twitter:** \$134,861

**Microsoft:** \$119,129

**LinkedIn:** \$138,798



**BUSINESS INSIDER CAREERS**

## Here's how much money you make in the 'sexiest job of the 21st century'

Jacquelyn Smith   

Feb. 25, 2016, 3:20 PM  33,395

[1] T.H. Davenport and D.J. Patil, 2012, Data Scientist: The Sexiest Job of the 21st Century, *Harvard Business Review* (Oct.)

<https://www.datasciencecentral.com/>

[https://www.glassdoor.com.br/Salaries/data-scientist-salary-SRCH\\_KO0,14.htm?countryRedirect=true](https://www.glassdoor.com.br/Salaries/data-scientist-salary-SRCH_KO0,14.htm?countryRedirect=true)

# O que se espera do Cientista de Dados?

- Teoria e Prática
  - Fundamentos de matemática e estatística
  - Banco de dados
  - Programação paralela e distribuída
  - Mineração de dados
  - Visualização
  - Conhecimento do domínio
  - Inovação
- Desenvolvimento de software
  - Diversas linguagens de programação
  - Diversas ferramentas



[1] H.B. Barua and K.C. Mondal, 2019, A comprehensive survey on cloud data mining (CDM) frameworks and algorithms, *ACM Computing Surveys*, v. 52, n. 5

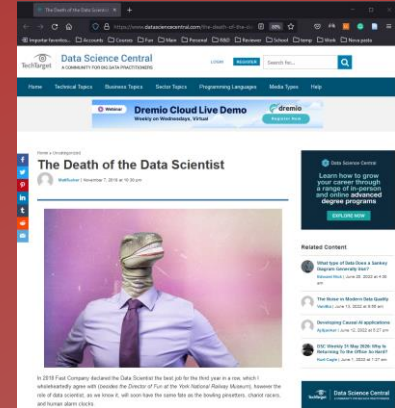
[2] D.A. Reed and J. Dongarra, 2015, Exascale computing and big data, *Communications of the ACM*, v. 58, n. 7, p. 56–68.

[3] V. Stodden, 2020, The data science life cycle, *Communications of the ACM*, v. 63, n. 7, p. 58–66.

# Ciência de Dados: A morte do Cientista de Dados

2018

- Existem um conjunto de atividades de Ciência de Dados que podem ser automatizadas
- Ferramentas vêm sendo desenvolvidas com boa usabilidade
  - ETL
  - Construção de Modelos
- Não é “apenas” produzir código básico em Python e R





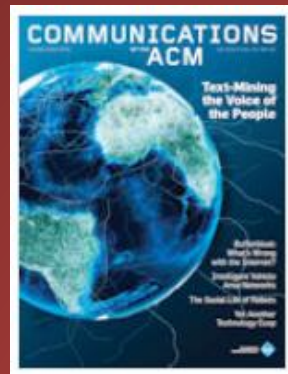
# Fundamentos...



ago/2002



dez/2008



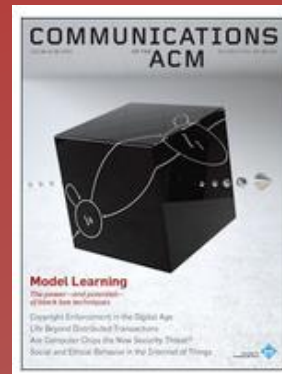
fev/2012



dez/2013



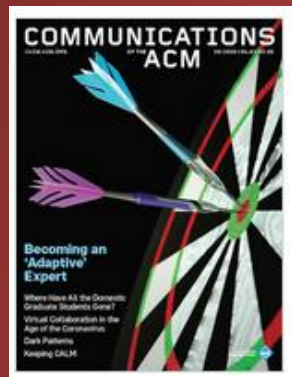
jul/2015



fev/2017



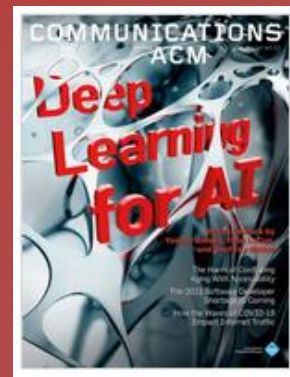
jul/2018



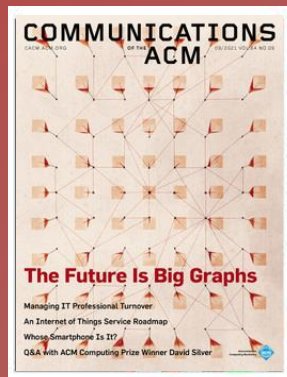
set/2020



mar/2021



jul/2021



sep/2021



mar/2022

[1] <https://cacm.acm.org/>

# Ciência de Dados: Cenário atual

- Nos EUA e em outros países as oportunidades são boas
  - Oportunidades vêm sem mantendo
  - Perspectivas é de alta até 2030
- No Brasil existe uma defasagem
  - Em 2014 havia poucas ofertas
  - Hoje o número vagas é razoável



*Changes in Base Salaries by Job Level for **Data Science Individual Contributors***

Job Level	Year	25%	Median	Mean	75%	N
Individual Contributor Level 1	2021	\$87,750	\$95,000	\$100,618	\$115,000	34
	2020	\$85,000	\$95,500	\$97,394	\$110,000	124
	Change	+3%	-0.5%	+3%	+5%	
Individual Contributor Level 2	2021	\$110,000	\$130,000	\$130,853	\$150,000	95
	2020	\$120,000	\$130,000	\$130,750	\$140,000	132
	Change	-8%	0%	0%	+7%	
Individual Contributor Level 3	2021	\$149,250	\$160,000	\$162,995	\$175,000	62
	2020	\$148,000	\$165,000	\$172,309	\$185,000	55
	Change	+1%	-3%	-5%	-5%	

# Perspectivas de Investimentos em TI no Brasil



## Perspectivas de Investimentos de 2021-2024 (R\$ bilhões)

### Tecnologias de Transformação Digital

### Mobilidade e Conectividade

Mobile, Dados e Banda Larga



R\$ 431,5 bi

4,6% a.a.



Taxa de câmbio: R\$/US\$ 3,95 (2019)



**Nuvem**  
R\$ 181,9 bi | 28% a.a.



**Robótica**  
R\$ 31,4 bi | 6% a.a.



**Internet das Coisas**  
R\$ 74,3 bi | 29% a.a.



**Realidade Virtual**  
R\$ 2,2 bi | 40% a.a.



**Big Data & Analytics**  
R\$ 77,2 bi | 13% a.a.



**Redes Sociais**  
R\$ 14,1 bi | 19% a.a.



**Segurança da Informação**  
R\$ 16,2 bi | 13% a.a.



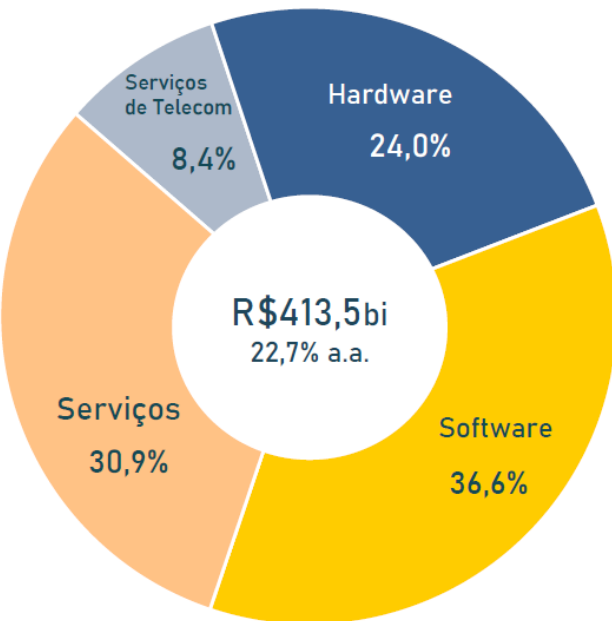
**Impressão 3D**  
R\$ 0,7 bi | 15% a.a.



**Inteligência Artificial**  
R\$ 11,5 bi | 31% a.a.



**Blockchain**  
R\$ 3,8 bi | 65% a.a.



# Carreira no Brasil

Entrar **glassdoor**

## Salários de Data Scientist Brasil

Visão geral **Salários** Entrevistas Plano de carreira

### Salários de Data Scientist em Brasil

Atualizado em 19 de jun. de 2022

**R\$ 8.000** /mês

Salário médio  
1.328 salários

R\$ 3 mil Baixa R\$ 14 mil Alta

Outras remunerações ⓘ  
Média: **R\$ 17.996** Intervalo: R\$ 1.617 - R\$ 42.749

Qual é o salário de Data Scientist? O salário médio nacional de Data Scientist é de R\$8.000 em Brasil. Filtre por localização para ver os salários de Data Scientist na sua região. As estimativas de salários têm como base os 1.328 salários enviados de forma sigilosa ao Glassdoor por funcionários com o cargo de Data Scientist.

Como a média salarial de R\$ 8.000 se aproxima da realidade para você?

😊 Correta   ↑ Alta   ↓ Baixa

Sua contribuição ajuda o Glassdoor a refinar nossas estimativas de salários com o tempo.

(7) Vagas de ciência de dados | x

https://www.linkedin.com/jobs/search/?keywords=ciência de dados

ciência de dados / Bra...

Vagas Data do anúncio Nível de experiência Todos os filtros

**ciência de dados em: Brasil**  
1.462 resultados Configurar alerta

- Engenharia de Machine Learning**  
Porto  
São Paulo, Brasil (Híbrido)  
17 ex-estudantes trabalham aqui  
Há 1 hora
- Cientista de Dados (remoto)**  
Data Science Brigade  
Porto Alegre, Rio Grande do Sul, Brasil (Remoto)  
Seu perfil corresponde a esta vaga  
Há 1 dia • Candidatura simplificada
- Cientista de Dados - Sênior**  
Mutant  
Rio de Janeiro, Brasil (Presencial)  
1 ex-estudante trabalha aqui  
Há 2 horas
- Pessoa Cientista de Dados**  
Zup Innovation  
Brasil (Remoto)  
1 conexão trabalha aqui  
Há 4 horas • Candidatura simplificada
- Cientista de Dados Sênior**  
banco BV  
São Paulo, São Paulo, Brasil  
1 conexão trabalha aqui  
Há 6 dias
- Staff Data Scientist**  
PicPay  
Brasil (Remoto)

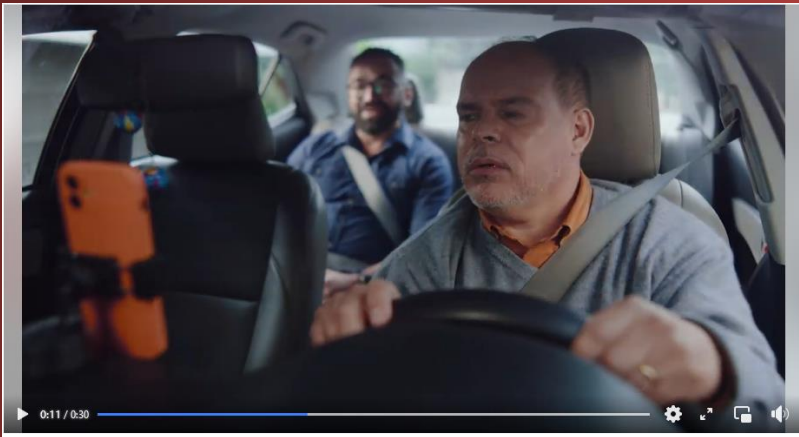
Mensagens



# Oportunidades

- Oferta aparece com várias nomenclaturas diferentes
  - Data Science
  - Machine Learning
  - Big Data, Business Intelligence, Data Analytics, Data Engineering

Propaganda do Itaú



**Não é uma panaceia**

*Estabilidade versus volatilidade*

*Expectativa versus realidade*

*É um nicho, portanto, deve ser encarado como um diferencial*

**Fatos e Dados**

## Concurso da Petrobras tem mais de 200 mil inscritos

13.jan.2022

Alta procura reflete interesse dos profissionais brasileiros em atuar na companhia. Há vagas para engenharías, geología, geofísica, **Ciência de dados**, entre outras carreiras

A Petrobras encerrou na quarta-feira (05/01/2022) as inscrições para o concurso público que selecionará 757 profissionais para cargos de nível superior. O número de inscritos no processo seletivo foi de 212 mil pessoas de todos os estados brasileiros. Levando-se em conta todas as carreiras, a relação candidato-vaga será de 566.

O elevado número de inscritos no concurso reflete o interesse do mercado de trabalho em atuar na companhia, que é reconhecida pela excelência de seu corpo técnico. Nos últimos anos, a Petrobras vem passando por um sólido processo de recuperação financeira. Por meio de uma gestão eficiente, a companhia vem melhorando expressivamente sua performance operacional e ampliando sua contribuição para a sociedade.

A prova está prevista para dia 20 de fevereiro e será realizada nas capitais de todos os estados do Brasil, além do Distrito Federal. Com o concurso, a Petrobras pretende selecionar profissionais com perfil dinâmico com foco em inovação e que tenham interesse em desenvolver suas habilidades em áreas estratégicas para a empresa.

**33 vagas**

**Visitas**  
39.388.924

**Categorias**

- Institucional
- O Negócio é o Seguinte
- Atividades
- Compliance
- Tecnologia e Inovação
- Sociedade e Meio Ambiente
- Produtos e Serviços
- Postos Petrobras

Ciência de dados



*Governo*

# Ética no governo

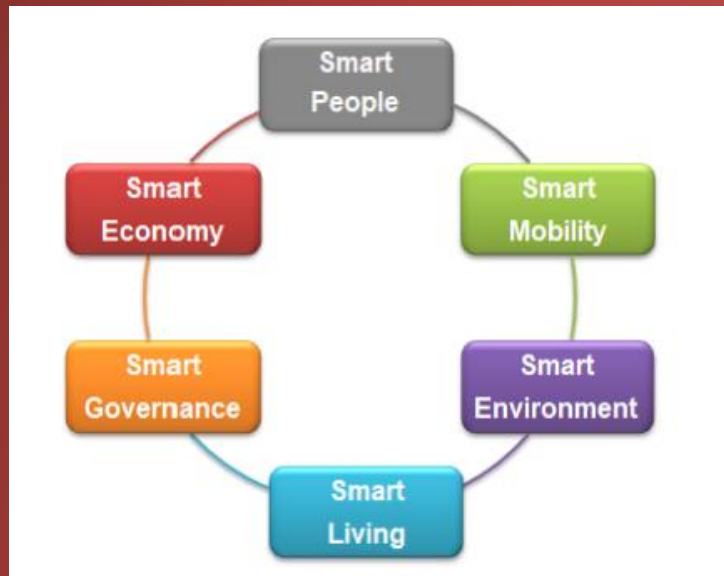
- Ciência de Dados pode oferecer insights aos políticos e tornar os serviços mais personalizados e eficientes
- Tratamento para os dados
  - Dados pessoais: identifica uma pessoa
  - Dados sensíveis: atenção no tratamento (criança/adolescente, racial/étnica, religiosa/filosófica, genética/saúde, partidária/sindical, sexual)
  - Dados públicos: finalidade, boa-fé, interesse público
  - Anonimizado: modificado (não pode ser reversível)
- Os países estão desenvolvendo leis para uso dos dados
  - <https://www.gov.br/cidadania/pt-br/aceso-a-informacao/lgpd>
  - Lei nº 13.709/2018
  - Papéis
    - Controlador
    - Operador
    - Encarregado
  - Compartilhamento na administração pública é previsto
- Modelos robustos



[1] C. Drew, 2016, Data science ethics in government, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, v. 374, n. 2083 <https://www.gov.br/cidadania/pt-br/aceso-a-informacao/lgpd>

# Cidades inteligentes

- Atores
  - Governantes, Empresas, Cientistas
- Fontes de Dados
  - IoT, sensoriamento populacional
- Serviços
  - Governo, saúde, educação, turismo
- Dimensões
  - Pessoas (dinâmica e comportamento), Mobilidade, Economia (atuação das empresas), Governança (serviços públicos), Vida (turismo, emergência), Ambiente (poluição)



[1] V. Moustaka, A. Vakali, and L.G. Anthopoulos, 2019, A systematic review for smart city data analytics, ACM Computing Surveys, v. 51, n. 5

# Iniciativas de Cidades Inteligentes



Rio  
PREFEITURA



## Prefeitura apresenta o Escritório de Dados e lança o Datalake

Publicado em 10/06/2022 - 14:23 | Atualizado em 10/06/2022 - 19:32

[Início](#) / [Cidade](#) / [Notícias](#) / Prefeitura apresenta o Escritório ...



Usamos cookies em nosso site para lhe dar a experiência mais relevante, lembrando suas preferências e repetindo visitas. Ao clicar em "Aceitar Tudo", você concorda com o uso de TODOS os cookies. No entanto, você pode visitar "Configurações de Cookies" para fornecer um consentimento controlado.

# Dados abertos

- Planejamento da gestão de dados abertos
- Transparência públicas



**PPCIC**  
PROGRAMA DE PÓS-GRADUAÇÃO  
EM CIÊNCIA DA COMPUTAÇÃO

O Programa de Pós-Graduação em Ciência da Computação (PPCIC) convida à



**AULA INAUGURAL**  
"Planejando o Futuro dos Nossos Dados – Planos de Gestão de Dados"

Prof. **Claudia Bauzer de Medeiros**  
UNICAMP

Data: 17/11/2020 – Horário: 17h  
Local: canal do YouTube do PPCIC



**PALESTRA**

**TRANSPARÊNCIA PÚBLICA  
E OS NOVOS LIMITES DA  
ABERTURA DE DADOS**



**GABRIELA RUBERG**  
BANCO CENTRAL DO BRASIL

**IX Workshop  
da EIC** 



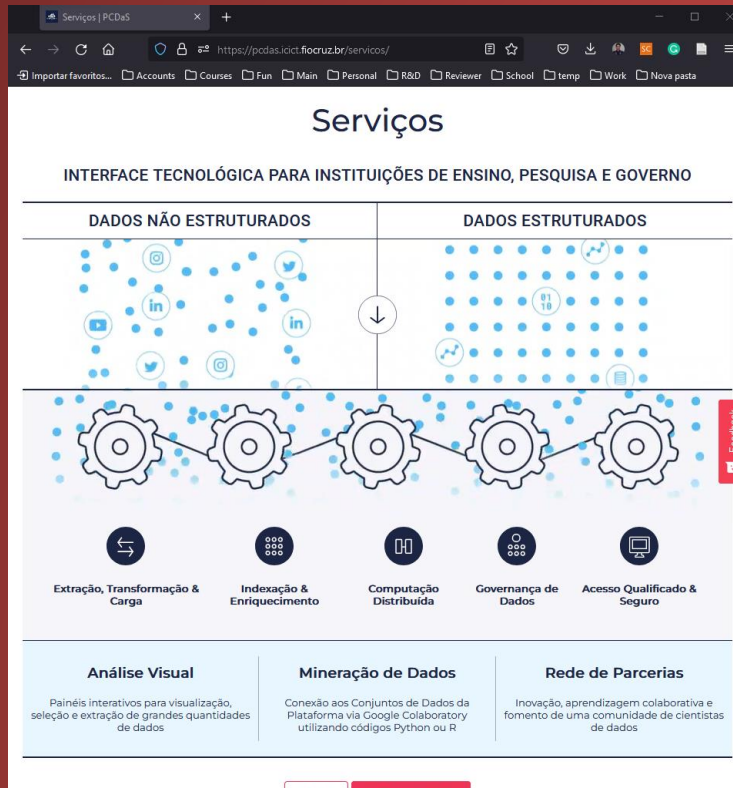
**20/10 • 14:00**

**PPCIC**  
PROGRAMA DE PÓS-GRADUAÇÃO  
EM CIÊNCIA DA COMPUTAÇÃO



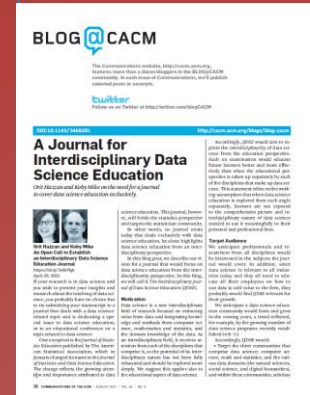
# Dados abertos e plataformas publicas para análise de dados

- A PCDaS é uma Plataforma pública, gratuita e em nuvem que disponibiliza grande quantidade de microdados e ferramentas de análise para pesquisas no campo da saúde



# Educação

- Atualmente há uma escassez de Cientista de Dados
  - EUA: Renovação dos programas existentes de Ciência de Dados ou criando novos, com um currículo vinculado ao setor
    - Aumento dos programas de ciência de dados de 13 em 2014 para 50 em setembro de 2020
- Necessidade de se discutir currículo
- ACM
  - Computing Competencies for Undergraduate Data Science Curricula (jan/2021)
- Temas abordados
  - Analysis and Presentation (AP)
  - Artificial Intelligence (AI)
  - Big Data Systems (BDS)
  - Computing and Computer Fundamentals (CCF)
  - Data Acquisition, Management, and Governance (DG)
  - Data Mining (DM)
  - Data Privacy, Security, Integrity, and Analysis for Security (DP)
  - Machine Learning (ML)
  - Professionalism (PR)
  - Programming, Data Structures, and Algorithms (PDA)
  - Software Development and Maintenance (SDM)

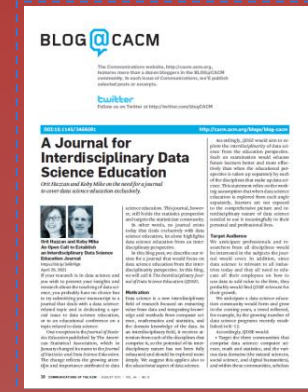


[1] O. Hazzan and K. Mike, 2021, A journal for interdisciplinary data science education, *Communications of the ACM*, v. 64, n. 8, p. 10–11.

<https://economictimes.indiatimes.com/nri/work/want-to-be-a-data-scientist-in-the-us-heres-what-you-need-to-know/articleshow/89017272.cms>

# Educação no Brasil

- INEP – Indicadores CINE (2021- Inclusão de área detalhada e rótulos na área geral 06- Computação e Tecnologias da Informação e Comunicação)
- Portaria SBC 20/2021 (jul/2021)
  - Graduação em Ciência de Dados
    - Angelo Brayner (UFC), Eduardo Ogasawara (CEFET/RJ), André de Carvalho (USP), Dulcan Ruiz (UFRGS)
  - Graduação em Inteligência Artificial



[1] O. Hazzan and K. Mike, 2021, A journal for interdisciplinary data science education, *Communications of the ACM*, v. 64, n. 8, p. 10–11.

<https://economictimes.indiatimes.com/nri/work/want-to-be-a-data-scientist-in-the-us-heres-what-you-need-to-know/articleshow/89017272.cms>

<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/cine-brasil/historico>

## *Eixos propostos para graduação*

- Fundamentos de Matemática, Estatística e Computação para Ciência de Dados
- Resolução de Problemas
- Desenvolvimento de Sistemas
- Engenharia e Exploração de Dados
- Gerência de Dados em Larga Escala
- Mineração de Dados e Aprendizado de Máquina
- Aprendizado Contínuo e Autônomo
- Ciência, Tecnologia, Inovação e Empreendedorismo

# *Análise da Pós-graduação em Ciência de Dados*

- Estudo realizado em 2015
  - Predominância nos EUA
  - Foram estudados mais de vinte programas
  - Organização e disciplinas oferecidas
- Área Interdisciplinar
  - Interação com diversos departamentos
- Sólida ligação com
  - Ciência da Computação
  - Estatística
  - Área de Aplicação (interface)



# CEFET/RJ – Programa de Pós-graduação em Ciência da Computação (PPCIC)

- Mestrado acadêmico criado desde 2016
  - Primeiro com o foco em Ciência de Dados
- Perfil do Egresso:
  - Sólida formação em Ciência da Computação
- Diferencial do Programa:
  - Formação de Cientistas de Dados
- Egresso deve estar apto a:
  - Atuar em setores onde há necessidade de extração de conhecimento a partir de dados
  - Tratar de questões teórico-práticas relacionadas aos problemas mais atuais relacionados à Computação
  - Propagar conhecimento no meio acadêmico-científico

# Disciplinas

## Básicas

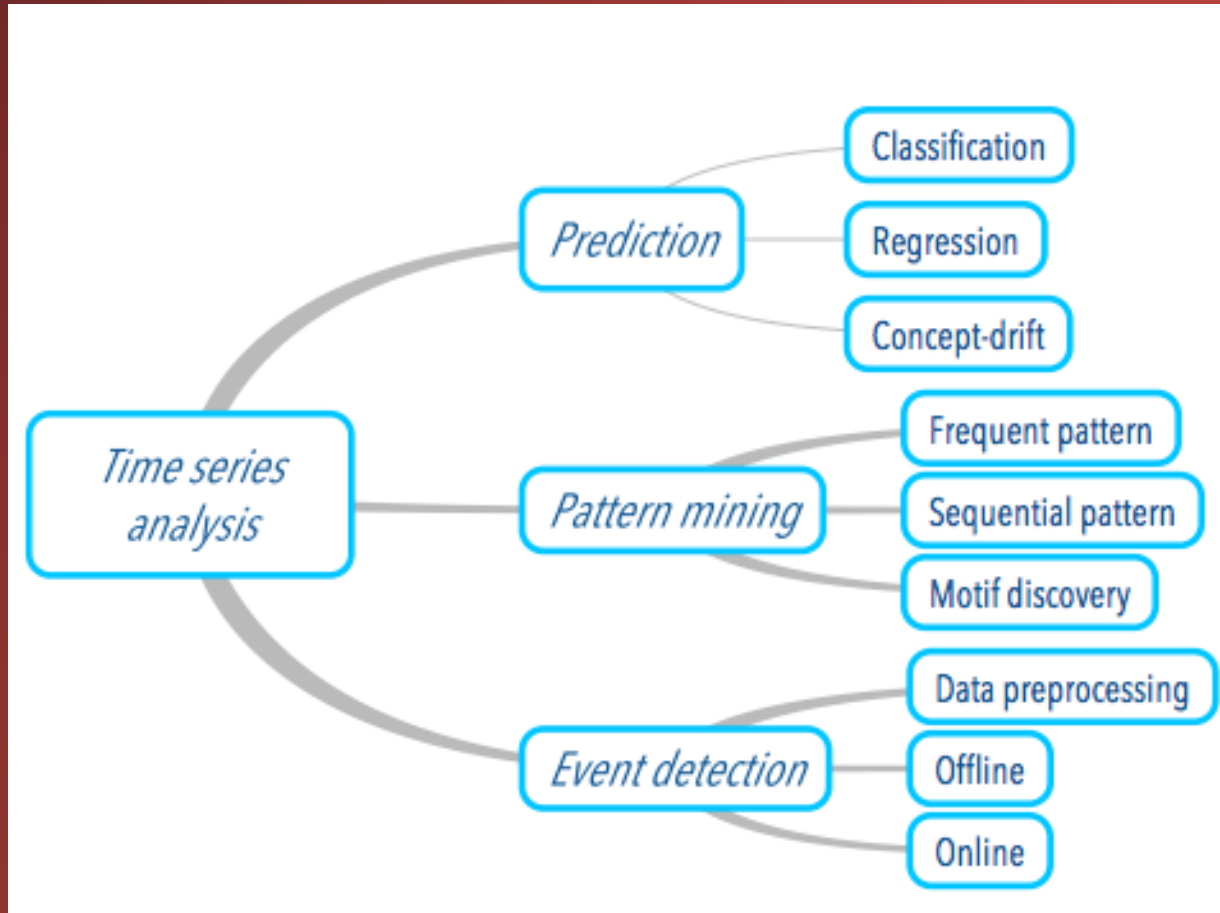
- Análise e Projeto de Algoritmos
- Arquitetura de Computadores
- Banco de Dados
- Computação Paralela e Distribuída
- Metodologia Científica em Computação
- Métodos Estatísticos

## Específicas

- Álgebra Linear Computacional
- Aprendizado de Máquina
- Ciência de Redes
- Gerência de Dados em Larga Escala
- Mineração de Dados
- Mineração de Textos
- Otimização por Metaheurísticas
- Visualização de Dados
- ...

*O que eu venho fazendo?*

# Data Analytics Lab



# Data Analytics Lab Team

The screenshot shows the DAL website with a navigation menu at the top: HOME, EQUIPE (with a dropdown arrow), CNPQ, APRESENTAÇÕES, ARTEFATOS, YOUTUBE, TUTORIAIS, and a search icon. The main content is organized into three sections:

- Doutorado**: Three team members are listed, each with a portrait photo and name/affiliation:
  - Lais Baroni (CEFET/RJ)
  - Leonardo Carvalho (CEFET/RJ)
  - Rebecca Salles (CEFET/RJ)
- Mestrado**: Six team members are listed, each with a portrait photo and name/affiliation:
  - Antônio Mello (CEFET/RJ)
  - Arthur Garcia (CEFET/RJ)
  - Cristiane Gea (CEFET/RJ)
  - Diego Salles (CEFET/RJ)
  - Janio Lima (CEFET/RJ)
  - Jéssica de Souza (CEFET/RJ)
- Graduação**: One team member is listed with a portrait photo and name/affiliation:
  - Baltazar Paixão (TCC)



# Novidades

YouTube BR Search SIGN IN

**Eduardo Ogasawara**  
445 subscribers

**SUBSCRIBE**

HOME VIDEOS PLAYLISTS CHANNELS ABOUT

Destaque ▶ PLAY ALL

**PESQUISAS em andamento**  
Eduardo Ogasawara  
249 views · 8 months ago

**Introdução ao R**  
Eduardo Ogasawara  
71 views · 3 weeks ago

**Metodologia Científica - Introdução**  
Eduardo Ogasawara  
237 views · 8 months ago

QR Code