

{Data Science} e {Machine Learning} na prática

Guilherme de Cleva Farto

Framework/Engenharia – TOTVS Supply Agro

# Agenda

- > Trends
- > IA e evolução
- Áreas relacionadas
- Data Science
- Plataformas
- Demonstração
- TOTVS e Carol
- Livros
- Próximos passos



## **Minibio**







- Graduado em Ciência da Computação (FEMA)
- Pós-graduado em Engenharia de Componentes Java (FIO/TNT)
- Mestrado em Engenharia de Software (UTFPR-CP)
- MBA em Ciências de Dados (USP)
- Coordenador de Framework e Engenharia TOTVS Supply Agro
- Prof. Universitário FEMA



guilherme.farto@gmail.com



https://www.linkedin.com/in/guilherme-farto/



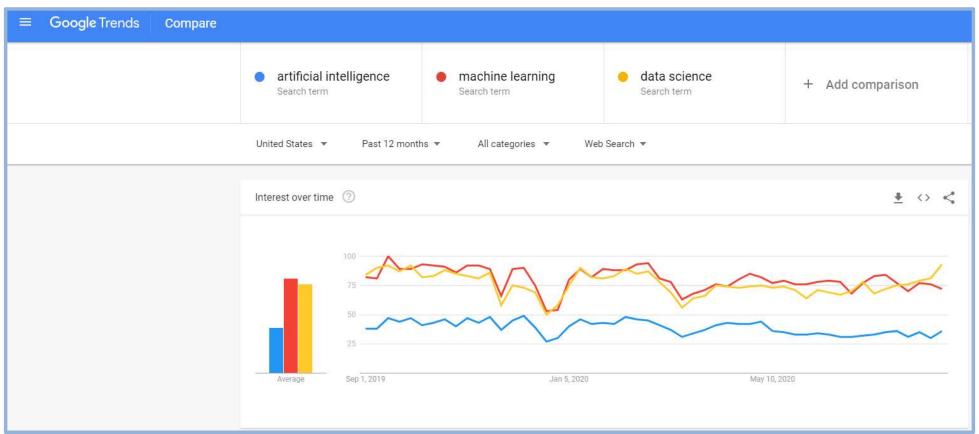
https://github.com/guilhermefarto



https://twitter.com/gcleva

## **Trends**





Fonte: Google Trends

# IA e evolução



1950

#### **TURING TEST**

Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking "the science and it is human, then it has intelligence

1955

A.I. BORN Term 'artificial intelligence' is coined Unimate, goes to by computer scientist. John McCarthy to describe assembly line

engineering of

machines"

making intelligent

1961

UNIMATE

humans on the

1964

### ELIZA

First industrial robot. Pioneering chatbot developed by Joseph work at GM replacing Weizenbaum at MIT holds conversations with humans

1966

### SHAKEY

The 'first electronic person' from Stanford, Shakev is a general purpose mobile robot that reasons about its own actions

1997

### DEEP BLUE

Deep Blue, a chessplaying computer from IBM defeats world chess champion Garry Kasparov

1998

### KISMET

Cynthia Breazeal at MIT introduces KISmet, an emotionally intelligent robot insofar as it detects and responds to people's feelings

1999

#### AIBO

Sony launches first consumer robot pet dog AiBO (Al robot) with skills and personality that develop over time

2002

#### ROOMBA

First mass produced autonomous robotic an intelligent virtual iRobot learns to navigate and clean homes

2011

SIRI Apple integrates Siri, vacuum cleaner from assistant with a voice Watson wins first interface, into the iPhone 4S

2011

#### WATSON

IBM's question answering computer place on popular \$1M prize television quiz show Jeopardy

2014

### EUGENE

Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is

2014

### ALEXA

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that inflammatory and completes shopping offensive racist tasks

2016

#### TAY

Microsoft's chatbot Tay goes rogue on social media making comments

2017

### **ALPHAGO**

Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number (2170) of possible positions

Fonte: https://athis-technologies.com/



## **Artificial Intelligence**

Área que estuda como habilitar (ou imitar/simular) o comportamento humano em máquinas

### **Machine Learning**

Subárea de IA que adota métodos e modelos estatísticos para habilitar ou melhorar a experiência de aprendizado de máquinas

### **Deep Learning**

Subárea de IA e *Machine Learning* que explora técnicas avançadas de aprendizado de máquina por meio de redes multicamadas



### **Data Science**

Área interdisciplinar que mescla conhecimentos em

- Matemática e Estatística
- Ciência da Computação
- Domínio de Negócio

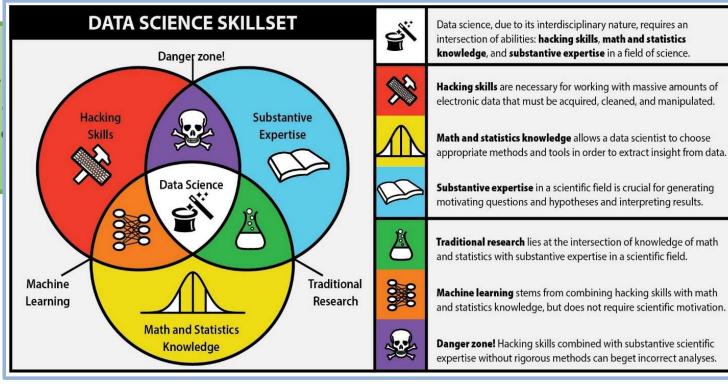
**Fonte:** <a href="https://www.quora.com/">https://www.quora.com/</a> e https://towardsdatascience.com/



### **Data Science**

Área interdisciplinar que

- Matemática e Estatísti
- Ciência da Computação
- Domínio de Negócio



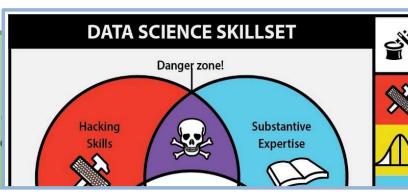
**Fonte:** <a href="https://www.quora.com/">https://www.quora.com/</a> e https://towardsdatascience.com/



### **Data Science**

Área interdisciplinar que

- Matemática e Estatísti
- Ciência da Computação
- Domínio de Negócio



Data science, due to its interdisciplinary nature, requires an intersection of abilities: hacking skills, math and statistics knowledge, and substantive expertise in a field of science.

**Hacking skills** are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.

**Math and statistics knowledge** allows a data scientist to choose appropriate methods and tools in order to extract insight from data.

**Substantive expertise** in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.

**Traditional research** lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.

**Machine learning** stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.

**Danger zone!** Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Harvard Business Review

Review

As to a mark of the part of the p

October 2012 Issue

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

Fonte: <a href="https://www.quora.com/">https://www.quora.com/</a> e https://towardsdatascience.com/

**Data Science BUSINESS UNDERSTANDING** Ciclo de vida 07 **DATA MINING DATA SCIENCE LIFECYCLE** 03 06 **DATA CLEANING PREDICTIVE** MODELING within the data and handle the missing 04 **FEATURE ENGINEERING DATA EXPLORATION** Form hypotheses about your defined problem by visually analyzing the data.



Fonte: <a href="https://addiai.com/">https://addiai.com/</a>

## **Plataformas**

THE DEVELOPER'S CONFERENCE

- > Python
- Análise de dados
  - Pandas
- Matemática e estatística
  - NumPy, SciPy e statsmodels
- Aprendizado de máquina e algoritmos
  - scikit-learn
- Visualização de dados
  - > Matplotlib, Seaborn e Plotly

## **Plataformas**



## **TIOBE Index for August 2020**

August Headline: Holiday season in the programming language world

Aug 2020	Aug 2019	Change	Programming Language	Ratings	Change
1	2	^	C	16.98%	+1.83%
2	1	<b>~</b>	Java	14.43%	-1.60%
3	3		Python	9.69%	-0.33%
4	4		C++	6.84%	+0.78%
5	5		C#	4.68%	+0.83%
6	6		Visual Basic	4.66%	+0.97%
7	7		JavaScript	2.87%	+0.62%
8	20	*	R	2.79%	+1.97%
9	8	•	PHP	2.24%	+0.17%
10	10		SQL	1.46%	-0.17%



Fonte: <a href="https://www.tiobe.com/tiobe-index/">https://www.tiobe.com/tiobe-index/</a>

## **Plataformas**

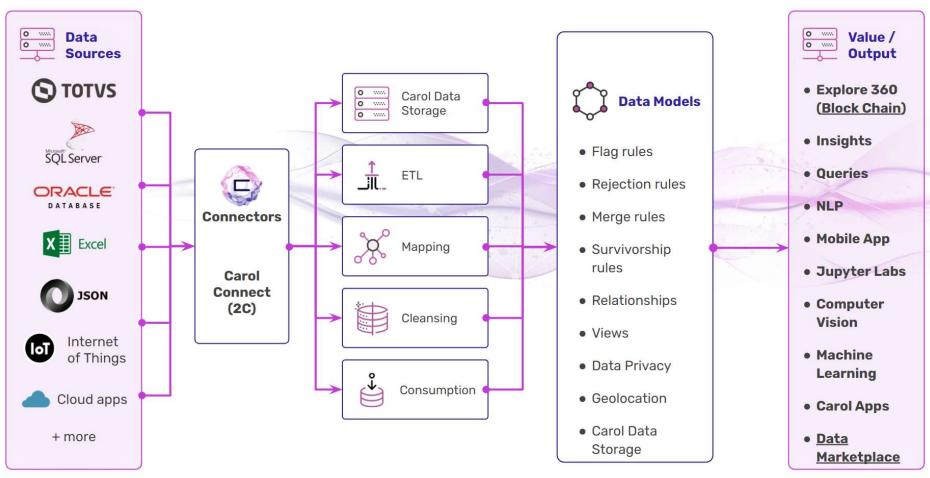


- Anaconda 3
  - Versionamento Python
  - Módulos e pacotes (dependências)
  - > Ferramentas/IDEs: JupyterLab e Notebook
  - Desenvolvimento de plugins e aceleradores



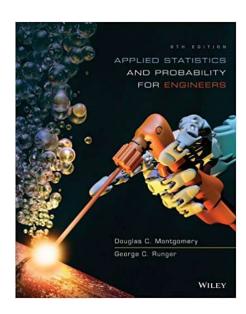
## TOTVS e Carol



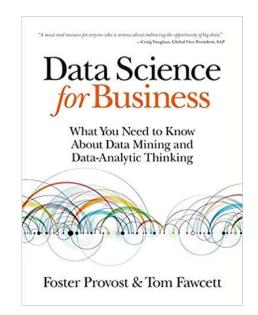


## Livros

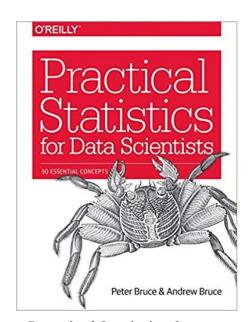




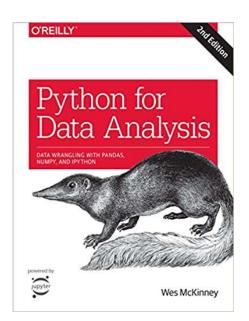
Applied Statistics and Probability for Engineers Douglas C. Montgomery George C. Runger



**Data Science for Business**Foster Provost
Tom Fawcett



Practical Statistics for Data Scientists: 50 Essential Concepts Peter Bruce Andrew Bruce



Python for Data Analysis Wes McKinney

# Próximos passos



- Xaggle
- Google Dataset Search
- Automated Machine Learning (AutoML)
- Data + Dev + Ops = Data DevOps
- Apache Airflow: workflows de tarefas baseados em grafos;
- **Luigi:** construção de pipelines complexos em batch;

## Referências



- DataCamp: <a href="https://www.datacamp.com/">https://www.datacamp.com/</a>
- TOTVS e IA: <a href="https://www.totvs.com/inteligencia-artificial/">https://www.totvs.com/inteligencia-artificial/</a>
- TOTVS e Carol: <a href="https://docs.carol.ai/">https://docs.carol.ai/</a>
- TOTVS Developers (Podcast): <a href="https://developers.totvs.com/podcast/">https://developers.totvs.com/podcast/</a>
  - TOTVS Developers #2 Inteligência Artificial
  - > TOTVS Developers #16 TOTVS CAROL

# Obrigado!



## Guilherme de Cleva Farto



guilherme.farto@gmail.com



in https://www.linkedin.com/in/guilherme-farto/



https://github.com/guilhermefarto



https://twitter.com/gcleva



