

# Inteligência Artificial e a era de Big Data em Astronomia

---

Lilianne Nakazono, PhD  
Observatório Nacional

Escola de Inverno em Astrofísica 2025

# Sobre mim

Bacharela em Estatística (IME-USP)

Doutora em Astronomia (IAG-USP)

- Doutorado sanduíche na University of Washington

Colaborações internacionais:

- S-PLUS
- J-PAS
- COIN

Coordenação do Grupo de Astroinformática

*Por favor não gravar ou tirar fotos em slides que tiverem esse ícone!*





## Grupos de Pesquisa



Astrofísica  
Computacional



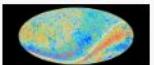
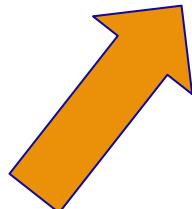
Astrofísica Estelar e  
Galáctica



Astrofísica Extragaláctica



Ciências Planetárias



Cosmologia e  
Astropartículas



# Conteúdo

**01**

## Introdução

O que é inteligência artificial?

**03**

## Fundamentos

Uma breve descrição matemática

**02**

## IA na Astronomia

Como se deu o crescimento de IA na Astronomia?

**04**

## Aplicações

Um pouco do meu trabalho

Conteúdos extras (se der tempo!): iniciativas brasileiras e ferramentas úteis

01

# Introdução

---



# Inteligência Artificial

O que vem à sua mente quando falamos de inteligência artificial?

## O que vem à sua mente quando falamos de Inteligência Artificial?

## Respostas dos participantes da Escola de Inverno do ON 2025:



# Inteligência Artificial

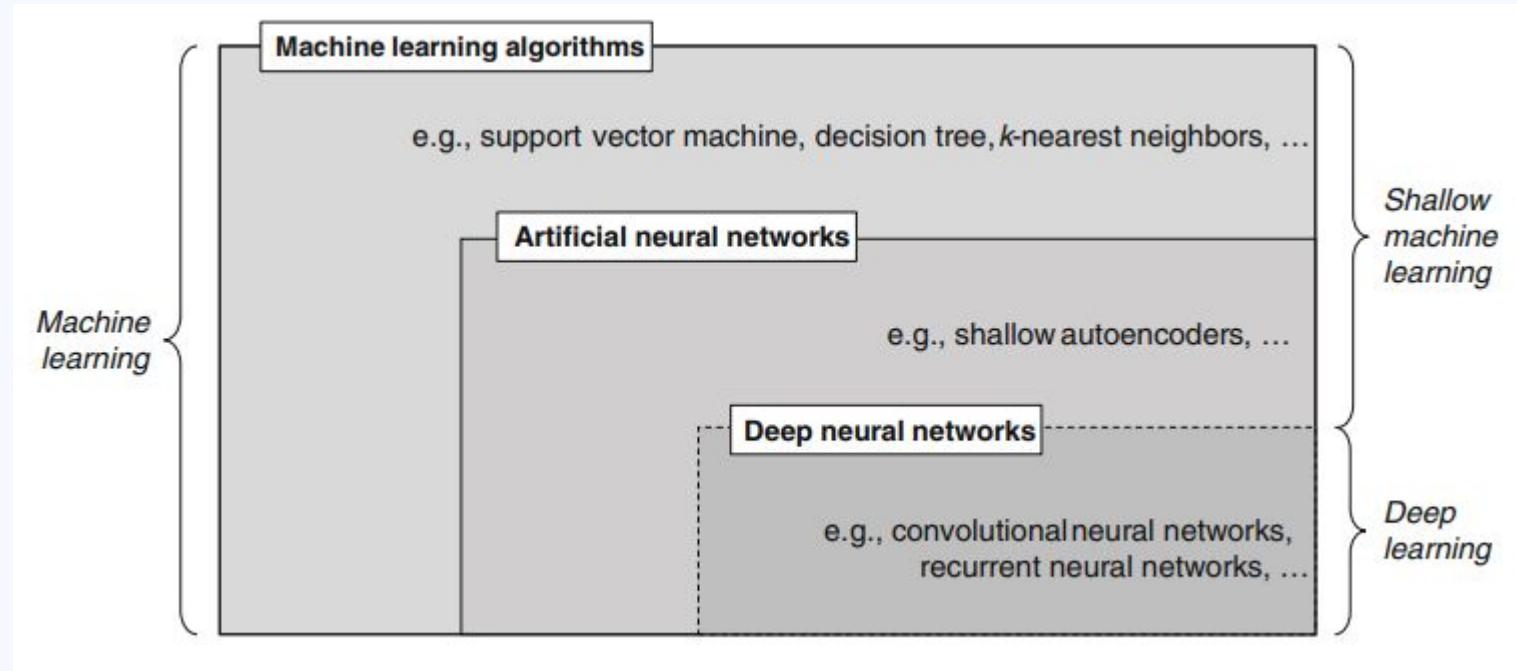


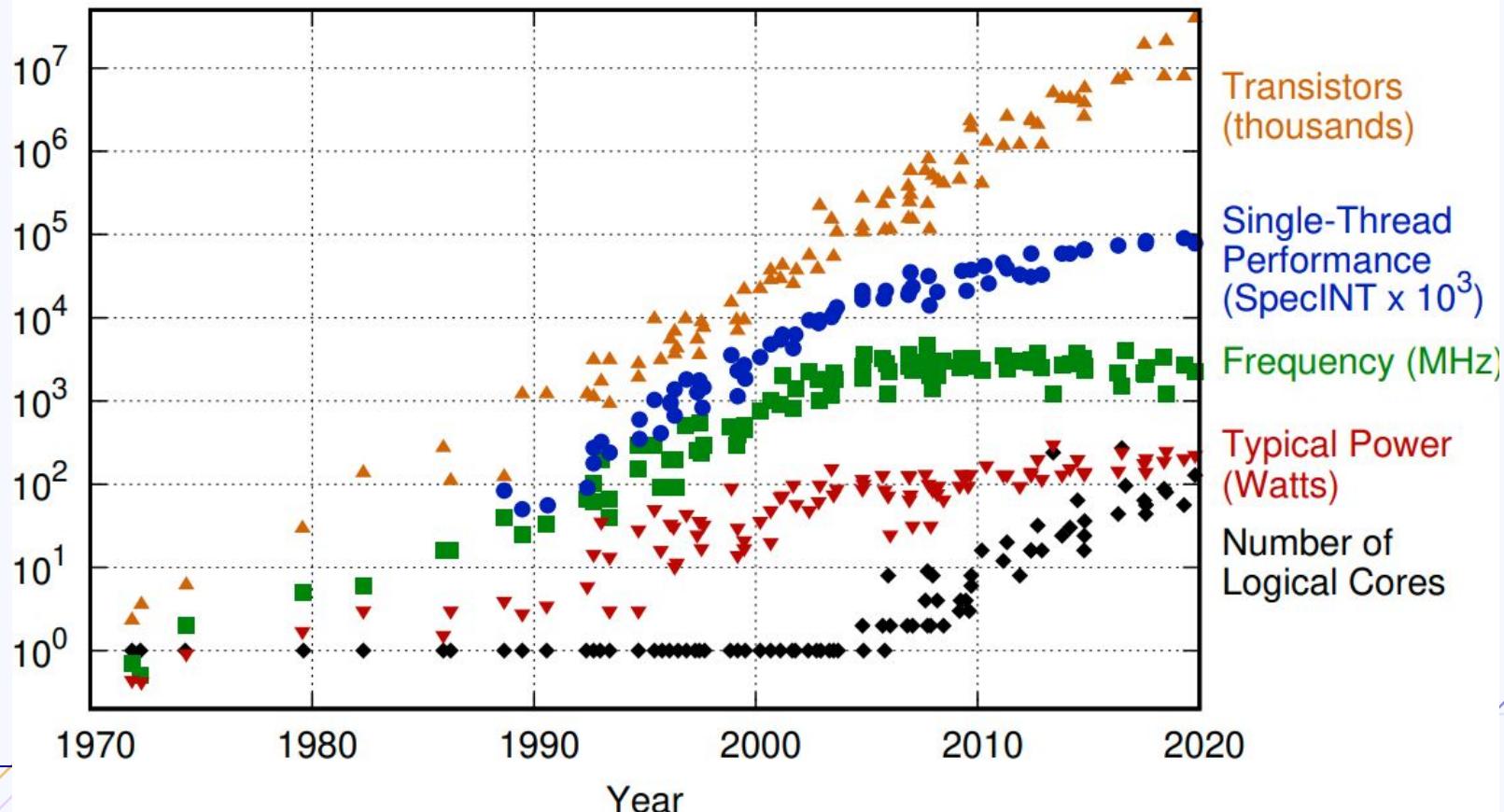
Diagram of Machine learning algorithms learning concepts and classes (inspired by Goodfellow et al. 2016, p. 9). Source: Janiesch, Zschech & Heinrich, 2021

# Não é um conceito novo!

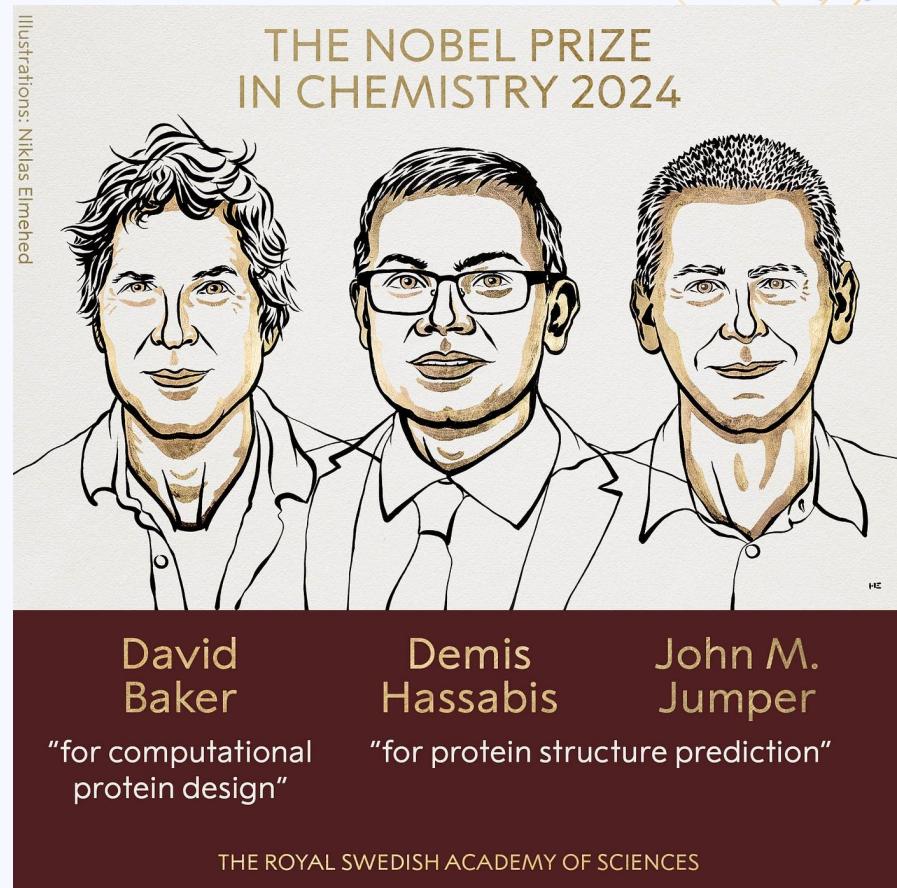
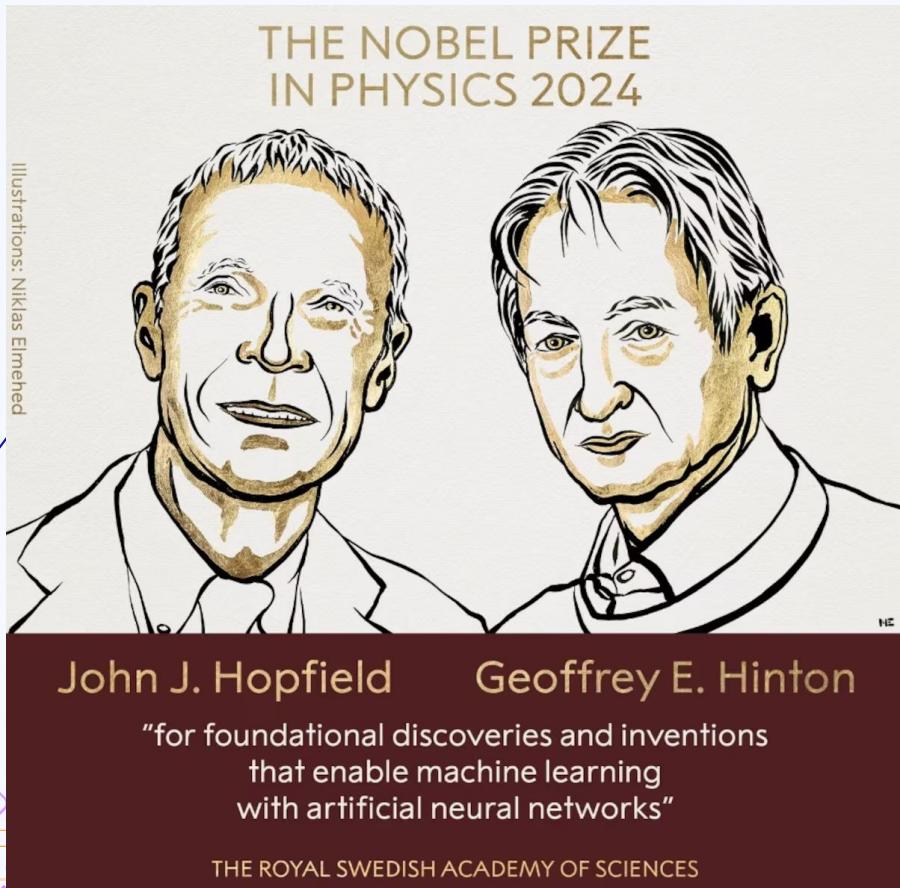
- O termo “Inteligência Artificial” surge em 1956 em uma conferência
- Proposição de redes neurais artificiais (Rosenblatt, 1957)
- Primeiro uso de redes neurais em artigo na Astronomia (Jeffrey & Rosner, 1986)

Por que o “boom” da IA só aconteceu recentemente?

# 48 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten  
New plot and data collected for 2010-2019 by K. Rupp



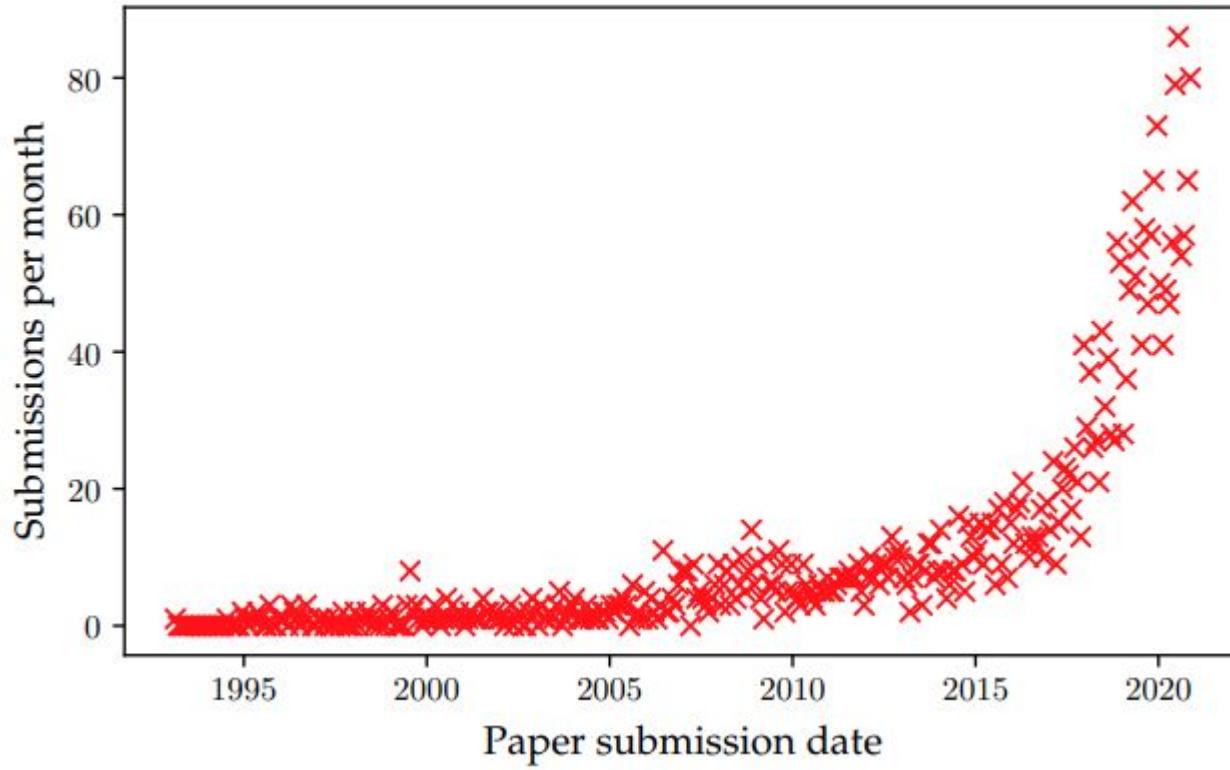
02

# IA na Astronomia

---

Um contexto histórico





Quantidade de submissões por MÊS no arXiv:astro-ph cujos resumos contêm os termos "machine learning", "ML", "artificial intelligence", "AI", "deep learning", ou "neural network" (Smith & Geach, 2023)

# IA na Astronomia

(Smith & Geach)



## 1<sup>a</sup> onda 1990~2010

Surgem os primeiros levantamentos do céu (ex: SDSS)



## 2<sup>a</sup> onda 2010~

Ascensão das redes neurais convolucionais e redes neurais recorrentes



## 3<sup>a</sup> onda hoje

Deep learning auto-supervisionado e generativo

# IA na Astronomia

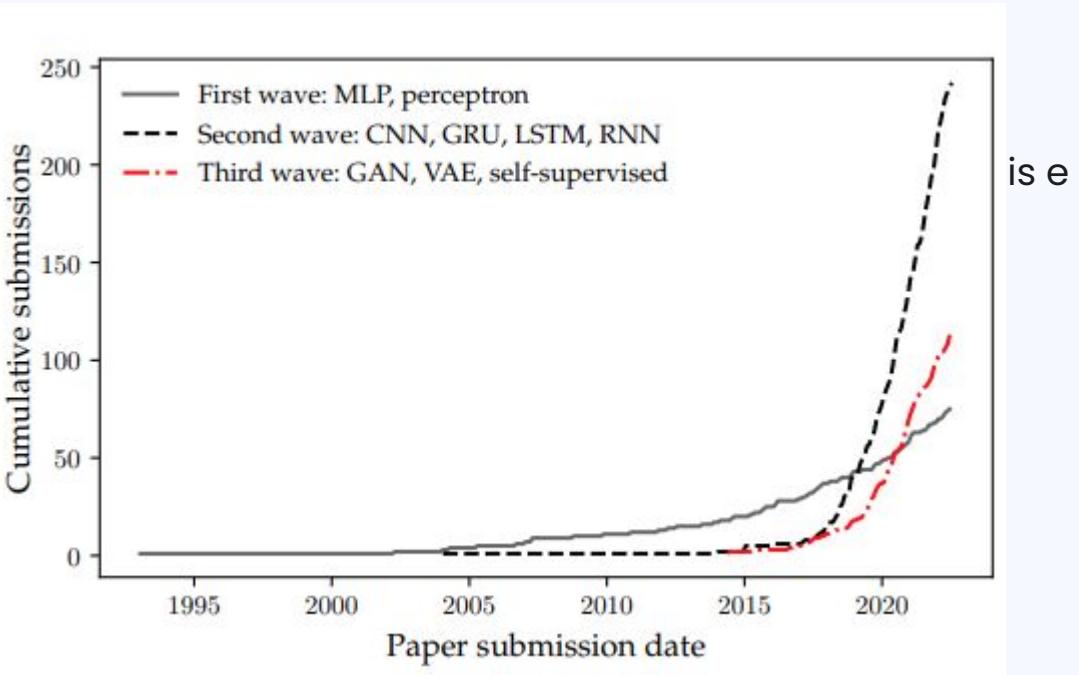
(Smith & Geach)

1<sup>a</sup> onda

Surgem levantado (ex: SDS)

3<sup>a</sup> onda

Deep learning, auto-supervised, generativa



# IA na Astronomia

(Smith & Geach)



## 1<sup>a</sup> onda 1990~2010

Surgem os primeiros levantamentos do céu (ex: SDSS)



## 2<sup>a</sup> onda 2010~

Ascensão das redes neurais convolucionais e redes neurais recorrentes



## 3<sup>a</sup> onda hoje

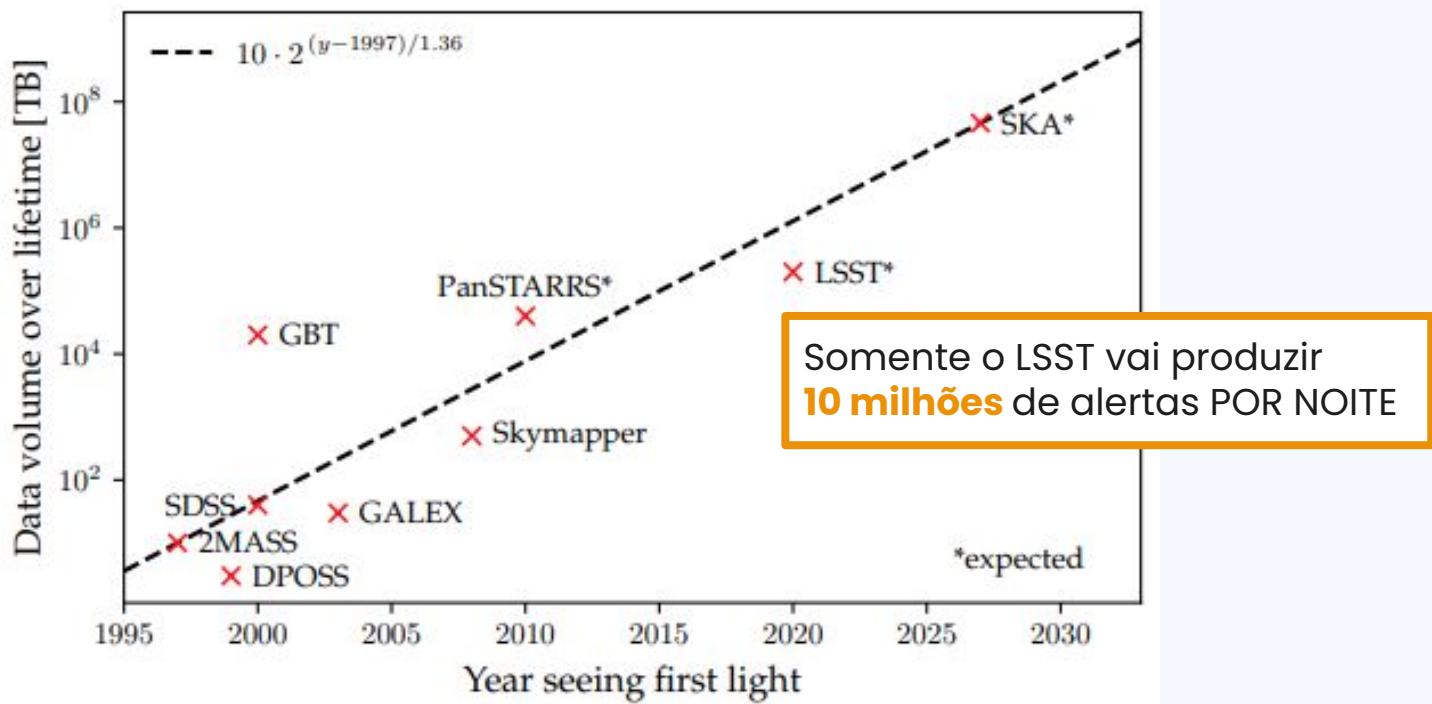
Deep learning auto-supervisionado e generativo



## ... próxima onda?

Foundation models (ex: GPT)

# E será que estamos preparados?



Volume total de dados de alguns levantamentos do céu (Smith & Geach, 2023)

# 03

# Fundamentos

---

Mostrando o fundamento matemático por trás do conceito de “aprendizado” algorítmico em poucos slides...



# Linear Regression

Consider an one-dimensional input and output, i.e.,  $(x^{(i)}, y^{(i)}) \in \mathbb{R}^2$ ,  $i = 1, 2, \dots, N$ .

We want to find a function  $f$  that relates each  $x^{(i)}$  to  $y^{(i)}$  in the **best way possible**:

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$

A family of hypothetical functions  $h: \mathbb{R} \rightarrow \mathbb{R}$  can be written as:

$$h_w(x) = w_0 + w_1 x$$

How can we find  $w_0$  and  $w_1$ ?

Answer: by minimizing a loss function  $\mathcal{L}$

$$\mathcal{L}(w_0, w_1) = \sum_{i=1}^N (w_0 + w_1 x^{(i)} - y^{(i)})^2$$

# Linear Regression: Analytical Solution (OLS)

Minimizing a loss function  $\mathcal{L}$  means taking its derivative:

$$\hat{w} = \arg \min_w \mathcal{L}(w_0, w_1)$$

For a 2-dimensional case:

$$\hat{w}_0 = \bar{Y} - w_1 \bar{X}$$

$$\hat{w}_1 = \frac{\sum_{i=1}^N (x^{(i)} - \bar{X})(y^{(i)} - \bar{Y})}{\sum_{i=1}^N (x^{(i)} - \bar{X})^2}$$

**Analytical solution** for any size of  $w$  (Ordinary Least Squares estimator):

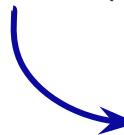
$$\hat{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

There are some advantages of having this analytical solution, but I won't dive into these details here!

# Numerical solution (Gradient descent)

Why do we care about a numerical solution for linear regression if we have an analytical one?

$$\hat{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$



inverting  $\mathbf{X}^T \mathbf{X}$  can be very computationally expensive! OLS works well for small datasets

Plus, OLS only works for linear models. If you have very large dataset, many features and wants to fit a nonlinear model, you need a numerical solution method.

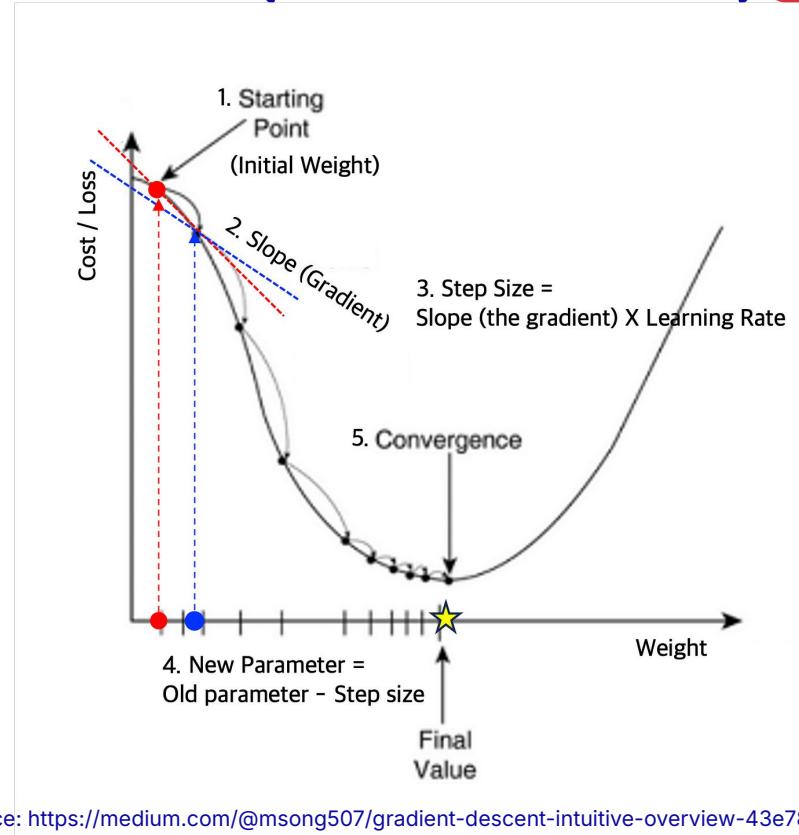
**This will be very important for neural networks!**

# Visualization of Gradient Descent (1-dimensional)

Imagine that the curve in this illustration is the loss function

$$\mathcal{L}(w_0, w_1) = \sum_{i=1}^N (w_0 + w_1 x^{(i)} - y^{(i)})^2$$

The idea of applying a numerical solution method is to find the minimum by iterative calculations



# Numerical solution (Gradient descent)

Iterative algorithm:

- Initialize parameters  $\mathbf{w}(0)$  with randoms
- Compute the gradient of  $\mathcal{L}$ , i.e.  $\nabla \mathcal{L}(\mathbf{w}(0))$
- Update the parameters  $\mathbf{w}$  as follows:

$$\mathbf{w}(1) = \mathbf{w}(0) - \eta \nabla \mathcal{L}(\mathbf{w}(0))$$

learning rate: arbitrary value

- Repeat until convergence

# Numerical solution (Gradient descent)

Iterative algorithm:

- Initialize parameters  $\mathbf{w}(0)$  with randoms
- Compute the gradient of  $\mathcal{L}$ , i.e.  $\nabla \mathcal{L}(\mathbf{w}(0))$
- Update the parameters  $\mathbf{w}$  as follows:

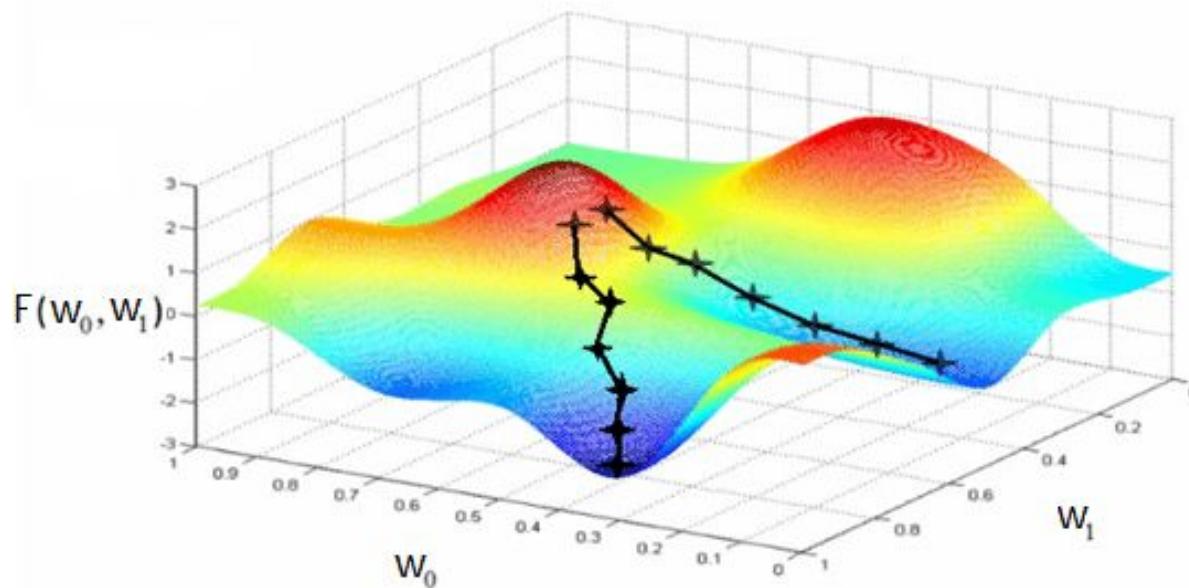
$$\mathbf{w}(1) = \mathbf{w}(0) - \eta \nabla \mathcal{L}(\mathbf{w}(0))$$

learning rate: arbitrary value

- Repeat until convergence



# Visualization of Gradient Descent (2-dimensional)



Source:

[https://www.researchgate.net/publication/322203555\\_Memristive\\_crossbar\\_arrays\\_for\\_machine\\_learning\\_systems](https://www.researchgate.net/publication/322203555_Memristive_crossbar_arrays_for_machine_learning_systems)

# Em resumo...

Algoritmos de machine learning e deep learning são, em essência:

- Minimização de uma função de perda
- Operações matriciais (e tensoriais) em alta dimensão



GPUs são ótimas pra isso!



# 04

# Aplicações

---

Um pouco do meu trabalho



# Tipos de problemas

Classificação

Régressão



# Tipos de problemas

## Classificação

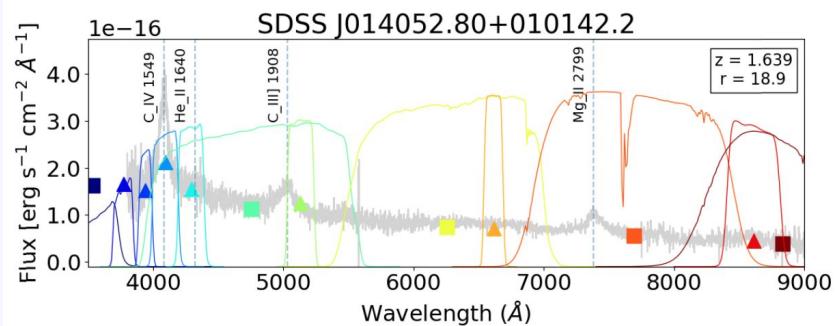
Exemplo: Que objeto celeste é este?



$Y = \{\text{star, galaxy, quasar, ...}\}$   
Valores discretos

## Regressão

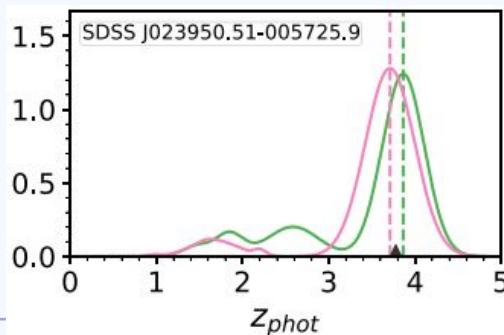
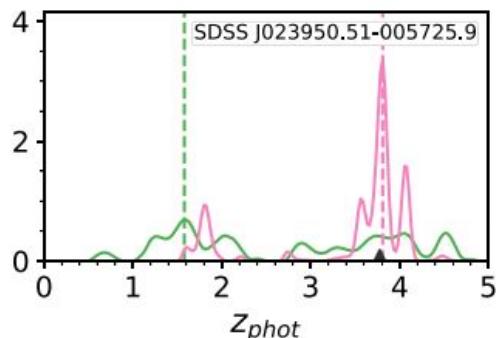
Exemplo: Conseguimos estimar redshift apenas com dados fotométricos?



$Y = \mathbb{R}$   
Valores contínuos

# Regressão

- Estimativa de redshift fotométrico de quasares usando redes neurais bayesianas e FlexCoDE



Monthly Notices  
of the  
ROYAL ASTRONOMICAL SOCIETY  
MNRAS 531, 327–339 (2024)  
Advance Access publication 2024 April 9  
<https://doi.org/10.1093/mnras/stac971>

**The Quasar Catalogue for S-PLUS DR4 (QuCatS) and the estimation of photometric redshifts**

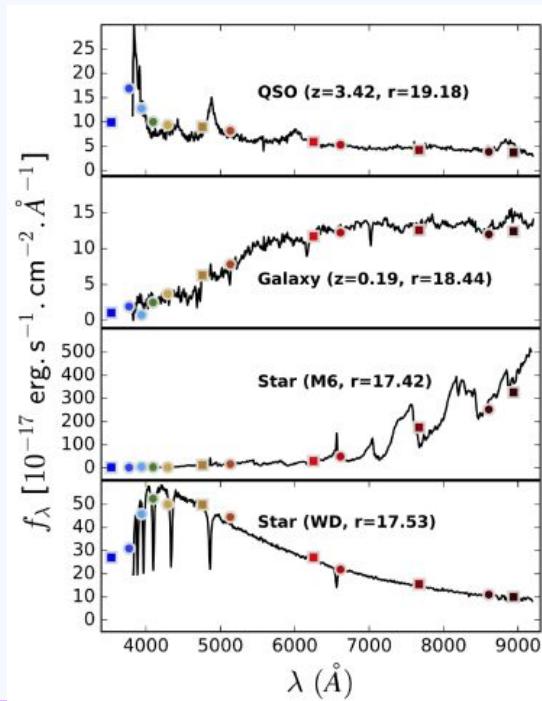
L. Nakazono   <sup>1,2</sup>★ R. R. Valença,<sup>1</sup> G. Soares,<sup>3</sup> R. Izbicki,<sup>3</sup> Ž. Ivezić   <sup>2</sup>, E. V. R. Lima,<sup>1</sup> N. S. T. Hirata,<sup>4</sup> L. Sodré Jr  <sup>1</sup>, R. Overzier,<sup>5</sup> F. Almeida-Fernandes,<sup>1,6</sup> G. B. Oliveira Schwarz,<sup>7</sup> W. Schoenell,<sup>8</sup> A. Kanaan,<sup>9</sup> T. Ribeiro<sup>10</sup> and C. Mendes de Oliveira<sup>1</sup>

<sup>1</sup>Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Departamento de Astronomia, São Paulo, SP 05508-090, Brazil  
<sup>2</sup>Department of Astronomy and DiRAC Institute, University of Washington, Box 351580, Seattle, WA 98195, USA  
<sup>3</sup>Universidade Federal de São Carlos, Departamento de Estatística, São Carlos, SP 13565-905, Brazil  
<sup>4</sup>Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Ciência da Computação, SP 05508-090, São Paulo, Brazil  
<sup>5</sup>Observatório Nacional / MCTIC, Rua General José Cristiano 77, Rio de Janeiro, RJ 20921-400, Brazil  
<sup>6</sup>NSF's NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA  
<sup>7</sup>Universidade Presbiteriana Mackenzie, R. da Consolação 930 - Consolação, São Paulo, SP 01302-907, Brazil  
<sup>8</sup>GMTO Corporation 465 N. Halsted Street, Suite 250 Pasadena, CA 91107, USA  
<sup>9</sup>Departamento de Física, Universidade Federal de Santa Catarina, Florianópolis, SC 88040-900, Brazil  
<sup>10</sup>Rubin Observatory Project Office, 950 N. Cherry Ave., Tucson, AZ 85719, USA

Accepted 2024 March 26. Received 2024 March 25; in original form 2023 September 27

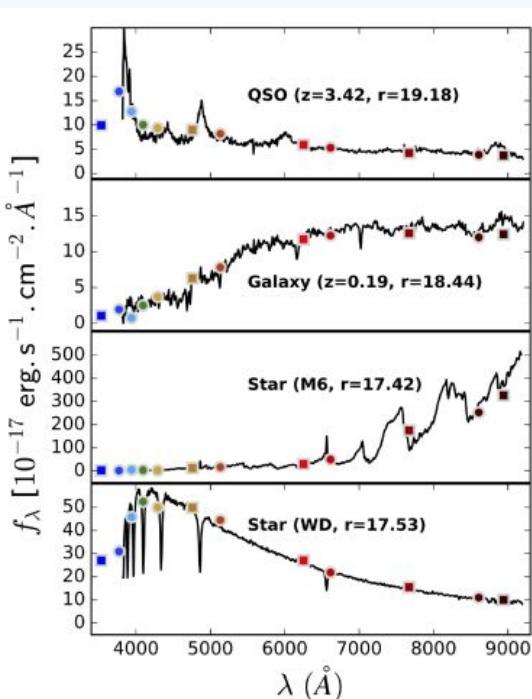
# Classificação

- Estrela/quasar/galáxia para o **S-PLUS** desde 2020 usando Random Forest



# Classificação

- Estrela/quasar/galáxia para o **S-PLUS** desde 2020 usando Random Forest



Monthly Notices  
of the  
ROYAL ASTRONOMICAL SOCIETY  
MNRAS 507, 5847–5868 (2021)  
Advance Access publication 2021 July 12  
<https://doi.org/10.1093/mnras/stab1835>

**On the discovery of stars, quasars, and galaxies in the Southern Hemisphere with S-PLUS DR2**

L. Nakazono <sup>1</sup>★ C. Mendes de Oliveira,<sup>1</sup> N. S. T. Hirata,<sup>2</sup> S. Jeram,<sup>3</sup> C. Queiroz <sup>4</sup>, Stephen S. Eikenberry,<sup>3</sup> A. H. Gonzalez <sup>3</sup> R. Abramo,<sup>4</sup> R. Overzier,<sup>5</sup> M. Espadoto,<sup>2</sup> A. Martinazzo,<sup>2</sup> L. Sampedro,<sup>1</sup> F. R. Herpich <sup>1</sup> F. Almeida-Fernandes,<sup>1</sup> A. Werle <sup>1,6</sup> C. E. Barbosa,<sup>1</sup> L. Sodré Jr. <sup>1</sup> E. V. Lima,<sup>1</sup> M. L. Buzzo,<sup>1</sup> A. Cortesi,<sup>7</sup> K. Menéndez-Delmestre,<sup>7</sup> S. Akras <sup>8</sup>, Alvaro Alvarez-Candal <sup>5,9,10</sup> A. R. Lopes,<sup>5</sup> E. Telles,<sup>5</sup> W. Schoenell,<sup>11</sup> A. Kanaan<sup>12</sup> and T. Ribeiro<sup>13</sup>

<sup>1</sup>Instituto de Astronomia, Geofísica e Ciências Atmosféricas da U. de São Paulo, Cidade Universitária, 05508-900 São Paulo, SP, Brazil  
<sup>2</sup>Departamento de Ciéncia da Computação, Instituto de Matemática e Estatística da USP, Cidade Universitária, 05508-090 São Paulo, SP, Brazil  
<sup>3</sup>Department of Astronomy, University of Florida, 211 Bryant Space Center, Gainesville, FL 32611, USA  
<sup>4</sup>Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, SP, Rua do Matão 1371 São Paulo, Brazil  
<sup>5</sup>Observatório Nacional / MCTIC, Rua General José Cristino 77, 20921-400 Rio de Janeiro, RJ, Brazil  
<sup>6</sup>INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Italy  
<sup>7</sup>Observatório do Valongo, Universidade Federal do Rio de Janeiro, Ladeira Pedro Antônio 43, Saíde, CEP 20080-090 Rio de Janeiro, RJ, Brasil  
<sup>8</sup>Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Penteli GR 15236, Greece  
<sup>9</sup>Instituto de Astrofísica de Andalucía, CSIC, Apt 3004, E-18080 Granada, Spain  
<sup>10</sup>IUFAcT, Universidad de Alicante, San Vicente del Raspeig, E-03080 Alicante, Spain  
<sup>11</sup>GMTI Corporation, 465 N. Halstead Street, Suite 250 Pasadena, CA 91107, USA  
<sup>12</sup>Departamento de Física, Universidade Federal de Santa Catarina, 88040-900 Florianópolis, SC, Brazil  
<sup>13</sup>NOAO, P.O. Box 26732, Tucson, AZ 85726, USA

Accepted 2021 June 18. Received 2021 June 16; in original form 2020 July 29

Relembrando um dos primeiros slides...

# IA na Astronomia

(Smith & Geach)



**1<sup>a</sup> onda** 1980~2000

Surgem os primeiros levantamentos do céu (ex: SDSS)



**2<sup>a</sup> onda** 2000~

Ascensão das redes neurais convolucionais e redes neurais recorrentes



**3<sup>a</sup> onda** hoje

Deep learning auto-supervisionado e generativo



**... próxima onda?**

Foundation models (ex: GPT)

# Classificação

Testamos se **Deep Learning** **auto-supervisionado** treinado com as imagens nos daria uma performance melhor que o modelo de Random Forest para o nosso problema

Experimento	Acurácia (%)	F-Score (%)
NW_RF	<b>96.87 <math>\pm</math>0.05</b>	<b>95.38 <math>\pm</math>0.13</b>
WW_RF	<b>98.44 <math>\pm</math>0.03</b>	<b>97.78 <math>\pm</math>0.05</b>
UF_RF	<b>98.43 <math>\pm</math>0.04</b>	<b>97.79 <math>\pm</math>0.08</b>
CNN	<b>95.45 <math>\pm</math>0.18</b>	<b>94.01 <math>\pm</math>0.24</b>

... e concluímos que não.

## Combinação de Dados Tabulares e Imagens para a Classificação de Objetos Astronômicos

G. Jacob Perin\*, L. Nakazono†, C. Mendes de Oliveira † e N. S. T. Hirata\*

\*Instituto de Matemática e Estatística

†Instituto de Astronomia, Geofísica e Ciências Atmosféricas

Universidade de São Paulo

**Resumo**—No contexto do S-PLUS (*Southern Photometric Local Universe Survey*), um projeto de imageramento do céu do hemisfério sul em 12 bandas fotométricas, foram desenvolvidos anteriormente métodos baseados em aprendizado de máquina para a classificação de três tipos de objetos (estrelas, galáxias e quasares). Um dos trabalhos utilizou dados de catálogo, incluindo objetos com dados faltantes, enquanto o outro utilizou imagens, sem objetos com dados faltantes. Neste trabalho apresentamos dois avanços: avaliamos os dois métodos em condições de igualdade e propomos a utilização de técnicas de *ensemble* para combinar os dois tipos de classificadores. Experimentos realizados com o quarto *Data Release* do S-PLUS mostram que o *ensemble* proposto supera ambos os métodos anteriores com respeito aos objetivos mais difíceis de serem classificados.

**Abstract**—In the context of S-PLUS (*Southern Photometric Local Universe Survey*), a 12-band photometric sky survey of the southern hemisphere sky, two machine learning based methods were previously developed for the classification of three types of objects (stars, galaxies and quasars). One of these works has used catalog data, including objects with missing information, while the other has used images, removing objects with missing information. In this work we present two advances: we evaluate the two methods under equal conditions and we propose the use of *ensemble* techniques to combine the two types of classifiers. Experiments performed with the fourth *Data Release* of the S-PLUS show that the proposed *ensemble* outperforms both previous methods with respect to the most difficult objects to be classified.

### I. INTRODUÇÃO

<https://doi.org/10.5753/sibgrapi.est.2023.27472>  
Conference on Graphics, Patterns and Images  
(SIBGRAPI)

porém utilizando uma seleção de dados mais conservadora e restritiva [3].

Neste trabalho, compararmos ambos os métodos propostos em condições de igualdade e propomos um *ensemble* que combina classificadores que utilizam dados tabulares e classificadores que utilizam imagens. Os resultados indicam que classificadores que utilizam dados de catálogo são superiores aos que utilizam imagens, e que o *ensemble* apresenta desempenho superior para os objetos mais difíceis de serem classificados do S-PLUS<sup>1</sup>.

Nas seções a seguir, explicamos inicialmente alguns conceitos de astronomia. Em seguida, descrevemos a metodologia adotada (seleção dos dados e treinamento dos modelos) e os resultados obtidos. Por fim, apresentamos as conclusões.

### II. CONCEITOS DE ASTRONOMIA

Diversos tipos de objetos podem ser observados no céu. Nesse trabalho, três tipos de objetos são considerados para a classificação: **estrelas, galáxias e quasares**.

Nesse contexto, duas formas de se realizar medições desses objetos são chamadas de **espectroscopia** - que gera um espectro relacionando fluxo luminoso com comprimento de onda - e **fotometria**, que consiste em capturar imagens de comprimentos de onda específicos. Enquanto a primeira carrega mais informações e é mais demorada de ser realizada, a segunda é mais rápida, porém torna a classificação dos objetos mais

# Deep Learning nem sempre é a melhor alternativa!



Às vezes é como matar uma formiga com uma bazuca!

Exige um bom conhecimento do problema e dos dados para escolher algoritmos ou métodos de análise

“Domain expertise” **sempre** vai ser necessário

# Conteúdo extra

## Iniciativas

---

Grupos de pesquisas interdisciplinares no Brasil com ênfase em Astronomia

[Documentation](#)[FAQ](#)[Tools](#)[Sign in](#)

# Astronomy Data Portal

## Smart System for Cosmic Exploration

The Astronomical Data Smart System (ADSS) is designed to maximize the efficiency of astronomical data distribution by integrating state-of-the-art technologies and artificial intelligence.

[Start Querying](#)[Explore Database](#)

[www.adss.cbpf.br](http://www.adss.cbpf.br)

Desenvolvido por um dos nossos  
membros (G. Schwarz)

**Advanced Astronomical Tools**



**AI Assisted Queries**

Leverage AI to retrieve objects without writing



**Data Visualization**

Interactive visualizations for complex



**Comprehensive Database**

Access millions of celestial objects with

[COIN](#)[Projects](#)[Residence Programs](#)[Focus](#)[Community](#)[Highlights](#)

# The Cosmostatistics Initiative



The Cosmostatistics Initiative (COIN) is a worldwide endeavor aimed to create an interdisciplinary community around data-driven problems in Astronomy.

It was designed to promote innovation in all aspects of academic scientific research.

The group is co-lead by [Rafael S. de Souza](#) (University of Hertfordshire, UK), [Emille E. O. Ishida](#) (Université Clermont Auvergne, France) and [Alberto Krone-Martins](#) (University of California Irvine, USA).

Researchers willing to join are welcome to contact any of the chairs!

<https://cosmostatistics-initiative.org/>

## 🐘 ELEPHANT: Extragalactic alErt Pipeline for Hostless AstroNomial Transients

P. J. Pessi<sup>1</sup>\*, R. Durgesh<sup>2</sup>, L. Nakazono<sup>3</sup>, E. E. Hayes<sup>4</sup>, R. A. P. Oliveira<sup>5</sup>, E. E. O. Ishida<sup>6</sup>,  
 A. Moitinho<sup>7</sup>, A. Krone-Martins<sup>8</sup>, B. Moews<sup>9,10</sup>, R. S. de Souza<sup>11</sup>, R. Beck<sup>12</sup>,  
 M. A. Kuhn<sup>11</sup>, K. Nowak<sup>11</sup>, and S. Vaughan<sup>13</sup> (for the COIN collaboration)

<sup>1</sup> The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova 106 91, Stockholm, Sweden

<sup>2</sup> Independent Researcher, Ingolstadt, Germany

<sup>3</sup> Instituto de Astronomia, Geofísica e Ciências Atmosféricas da USP, 05508-900, São Paulo, Brazil

<sup>4</sup> Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK

<sup>5</sup> Astronomical Observatory, University of Warsaw, Al. Ujazdowskie 4, 00-478 Warszawa, Poland

<sup>6</sup> Université Clermont Auvergne, CNRS/IN2P3, LPC, F-63000 Clermont-Ferrand, France

<sup>7</sup> CENTRA, Universidade de Lisboa, FCUL, Campo Grande, Edif. C8, 1749-016 Lisboa, Portugal

<sup>8</sup> Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA 92697, USA

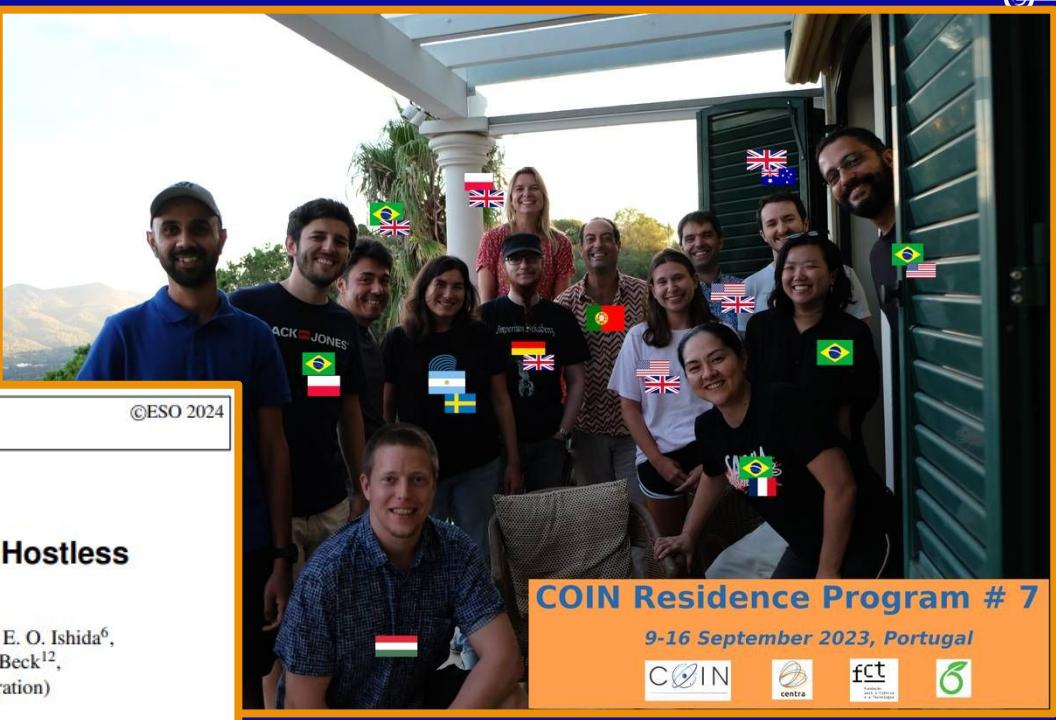
<sup>9</sup> Business School, University of Edinburgh, 29 Buccleuch Pl, Edinburgh, EH8 9JS, UK

<sup>10</sup> Centre for Statistics, University of Edinburgh, Peter Guthrie Tait Rd, Edinburgh, EH9 3FD, UK

<sup>11</sup> Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK

<sup>12</sup> Independent Researcher, Budapest, Hungary

<sup>13</sup> School of Mathematical and Physical Sciences, Macquarie University, NSW 2109, Australia



Curso de IA em Astronomia e Geofísica  
Quando: início de 2026

Fiquem de olho nos anúncios :)

# CIAC-ON

Esfôrço da diretoria, da divisão de tecnologia da informação e de pesquisadores da Astronomia e Geofísica computacional para a criação de um **centro de inteligência artificial do Observatório Nacional**



# Conteúdo extra

# Ferramentas úteis

---

# Programming Languages



python<sup>TM</sup>

Up to date, it is the most used programming language for data science in Astronomy



In Astronomy, we use **ADQL** (*Astronomical Data Query Language*) to query structured data. Syntax is very similar to SQL

# Python packages

Astronomy-specific



Data Visualization



Data manipulation



Mathematical computing



NumPy



SciPy

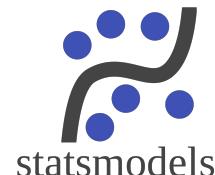
Image manipulation



Machine Learning



Statistical Modeling



Deep Learning



# Python packages

## Astronomy-specific



For distributed and parallel computing:  
(essential for dealing with big data!)



## Data manipulation

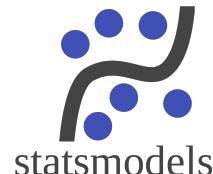


Mathematical

## Machine Learning



## Statistical Modeling



## Image manipulation



# Tools: code version control



Web-based platforms that hosts Git repositories:



# Tools: code editor

My personal recommendation:



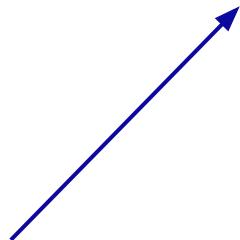
Visual Studio Code

With these extensions installed: Python, Jupyter, Remote - SSH, GitHub Copilot (plus others... Those make your programming life much much easier, trust me!)

# Programming best practices

Your data science project starts with setting up code environment (e.g. pyenv, conda) and properly organizing your code files!

One of my repositories

A screenshot of a GitHub repository interface. The repository structure is as follows:

```
> config  
> data  
> img  
> logs  
└ src  
    > __pycache__  
    > evaluation  
    > experiments  
    > models  
    > notebooks  
    > preprocess  
    > production  
    > scripts  
    > utils  
    &lt; __init__.py  
    &lt; .gitignore  
    ! environment.yml  
    &lt; LICENSE  
    &lt; README.md  
    &lt; setup.py  
    &lt; stilts.jar
```

The files listed include \_\_init\_\_.py, .gitignore, environment.yml, LICENSE, README.md, setup.py, and stilts.jar. The README.md file is marked with an info icon, and the stilts.jar file is marked with a red square icon.

# Obrigada !

## Perguntas?

Email: [liliannenakazono@on.br](mailto:liliannenakazono@on.br)

Website: [marixko.github.io](https://marixko.github.io)

GitHub: <https://github.com/marixko>

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons, infographics & images by [Freepik](#)