

Introdução ao Aprendizado de Máquina para Físicos

Marcelo Vargas dos Santos
Aula 1

Informações sobre o curso

Objetivo: Este curso é uma introdução ao uso de técnicas de Aprendizado de Máquina (Machine Learning, ML, em inglês) e suas aplicações na Física. O foco será nas aplicações e não nos métodos.

Horário: Quarta-Feira 10h e Quinta-Feira 14h

Sala: 105

Avaliação: Projeto

Ferramentas: Scikit-Learn, Keras (Python)

Conteúdo

Parte 1 - Aprendizado Estatístico (Statistical Learning)

1. Introdução: Big Data, Ciência de Dados e Aprendizado de Máquina.
2. Análise exploratória de dados.
3. Modelos Supervisionados com Scikit-Learn.
4. Otimização de Parâmetros.
5. Teste de Validação de modelos.
6. Pré-processamento de dados.
7. Modelos não supervisionados.

Parte 2 - Aprendizado Profundo (Deep Learning)

1. Introdução: O que são redes Neurais profundas.
2. Tensorflow e Keras.
3. Treinamento de redes neurais.
4. Redes Neurais Convolucionais. Inception, Resnet, InceptionNet.
5. Redes Neurais Recorrentes. LSTM.
6. Codificador automáticos e Redes adversárias generativas.
7. Aprendizado reforçado.



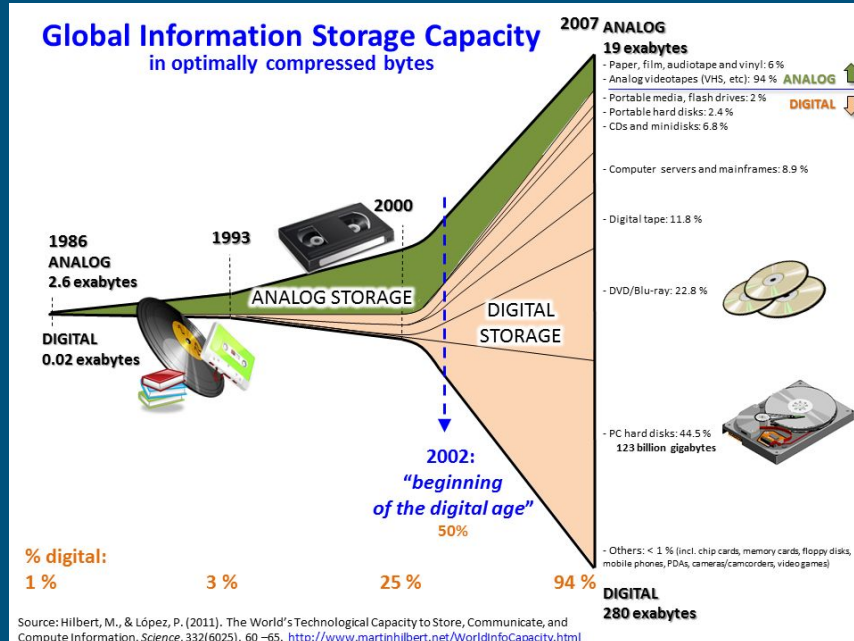
Motivação



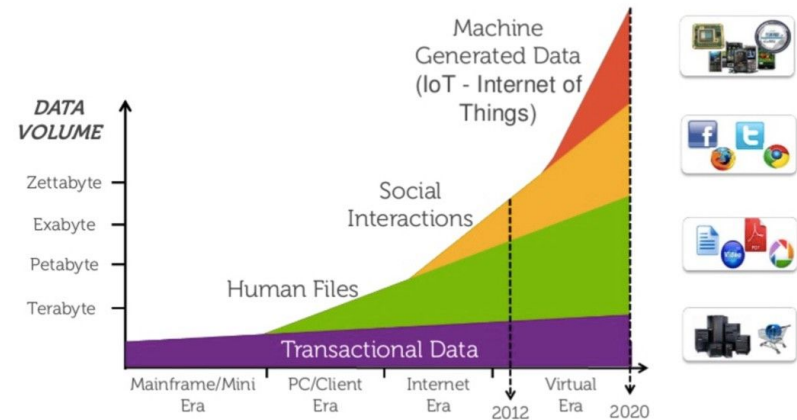
Big Data & Data Science



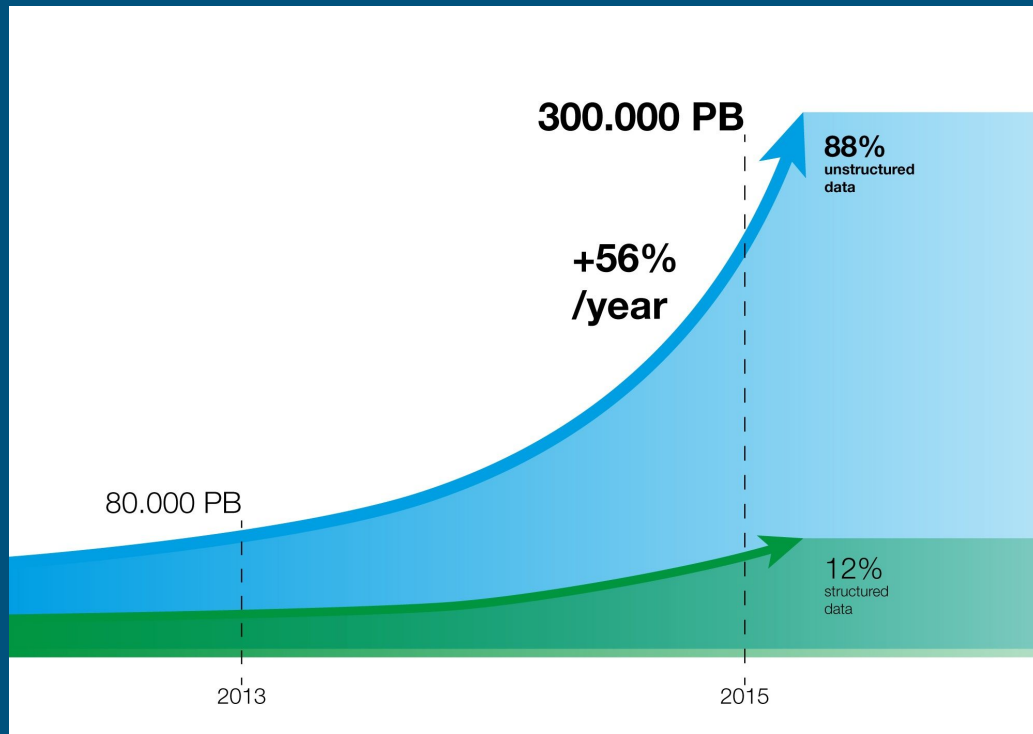
A era do Big Data



The Explosion of Data



Conhecido vs. Desconhecido



Dados científicos

Astronomia:

Sky Survey Projects	Data Volume
DPOSS (The Palomar Digital Sky Survey)	3 TB
2MASS (The Two Micron All-Sky Survey)	10 TB
GBT (Green Bank Telescope)	20 PB
GALEX (The Galaxy Evolution Explorer)	30 TB
SDSS (The Sloan Digital Sky Survey)	40 TB
SkyMapper Southern Sky Survey	500 TB
PanSTARRS (The Panoramic Survey Telescope and Rapid Response System)	~ 40 PB expected
LSST (The Large Synoptic Survey Telescope)	~ 200 PB expected
SKA (The Square Kilometer Array)	~ 4.6 EB expected

Table 1: Data volumes of different sky survey projects.







Outros campos: Altas Energias (LHC), Genética, Climatologia, Saúde Pública, Sociologia, etc.



0 Paradigma BIG DATA

The six Vs of big data

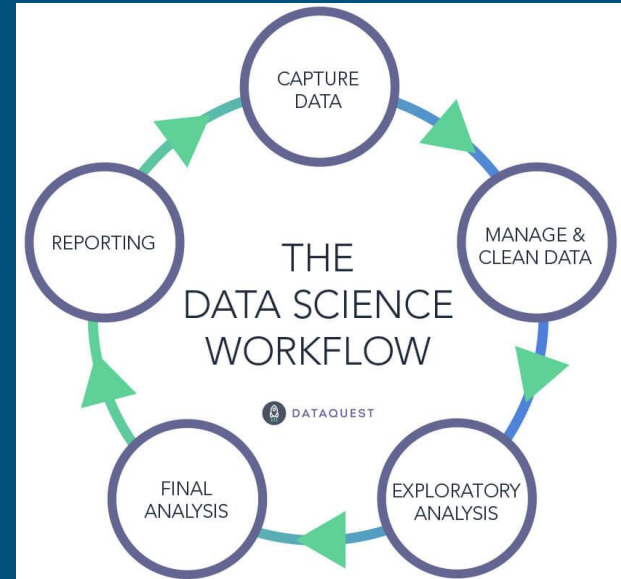
Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume, variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					

Ciência de Dados

É um campo interdisciplinar que usa métodos científicos, algoritmos e sistemas computacionais para extrair informação de dados estruturados e não estruturados.

Ferramentas: Programação, Estatística, Inteligência Artificial (Aprendizado de máquina).

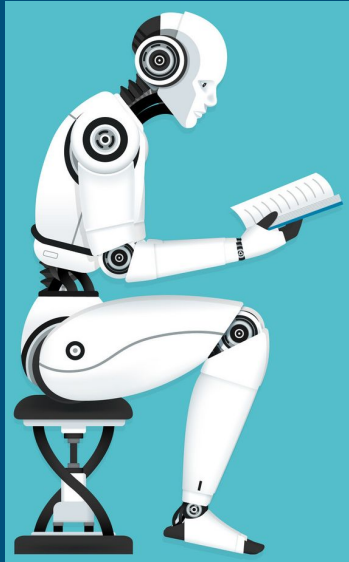




Machine Learning



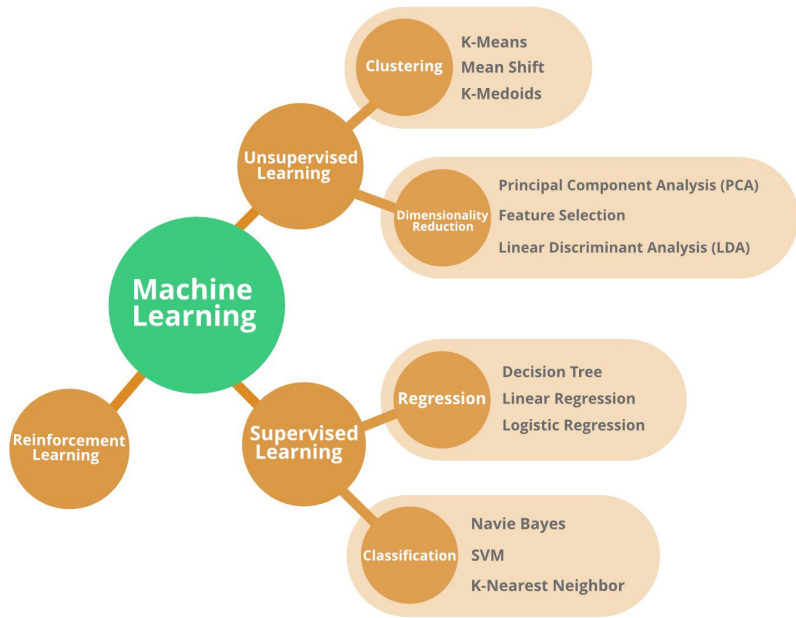
O que é Aprendizado de máquina?



Aprendizado de máquina (machine learning em inglês, ML) é o subcampo da Inteligência Artificial que estuda algoritmos e métodos estatísticos que sistemas computacionais usam para executar tarefas específicas sem uma programação explícita.

Exemplos: Reconhecimento de padrões, modelos preditivos, classificação e regressão.

Modelos de ML



Aprendizado supervisionado: Relacionam um conjunto de variáveis chamadas características (features), com um outro conjunto alvo (targets).

Aprendizado não supervisionado: Buscam relações entre variáveis não explicitamente relacionadas.

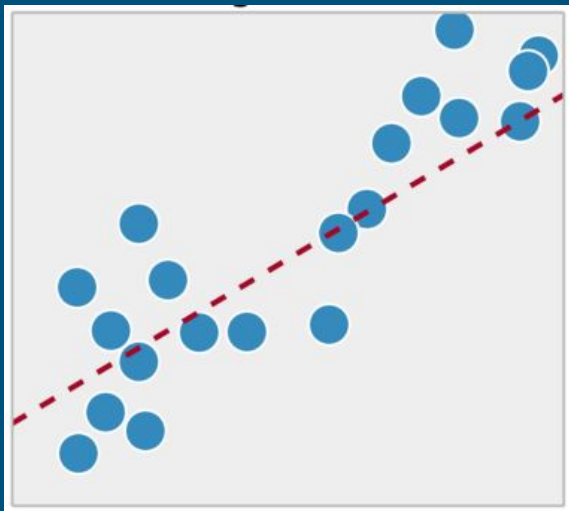
Reforçado: O modelo interage com um sistema a fim de coletar informação.

Misto: Combina duas das, ou as, 3 estratégias acima.

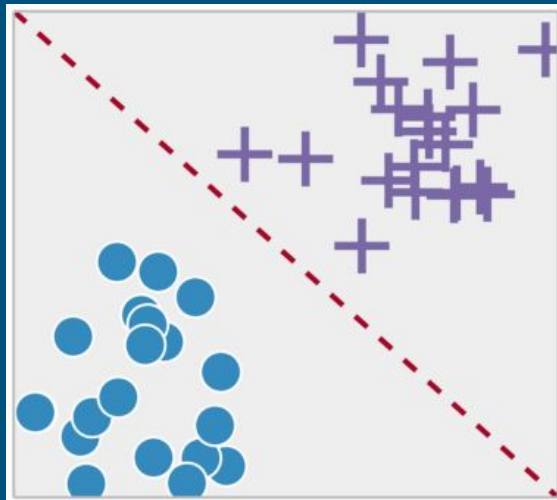
Aprendizado Supervisionado

Modelos preditivos para variáveis com rótulos

Rótulo Contínuo: Regressão

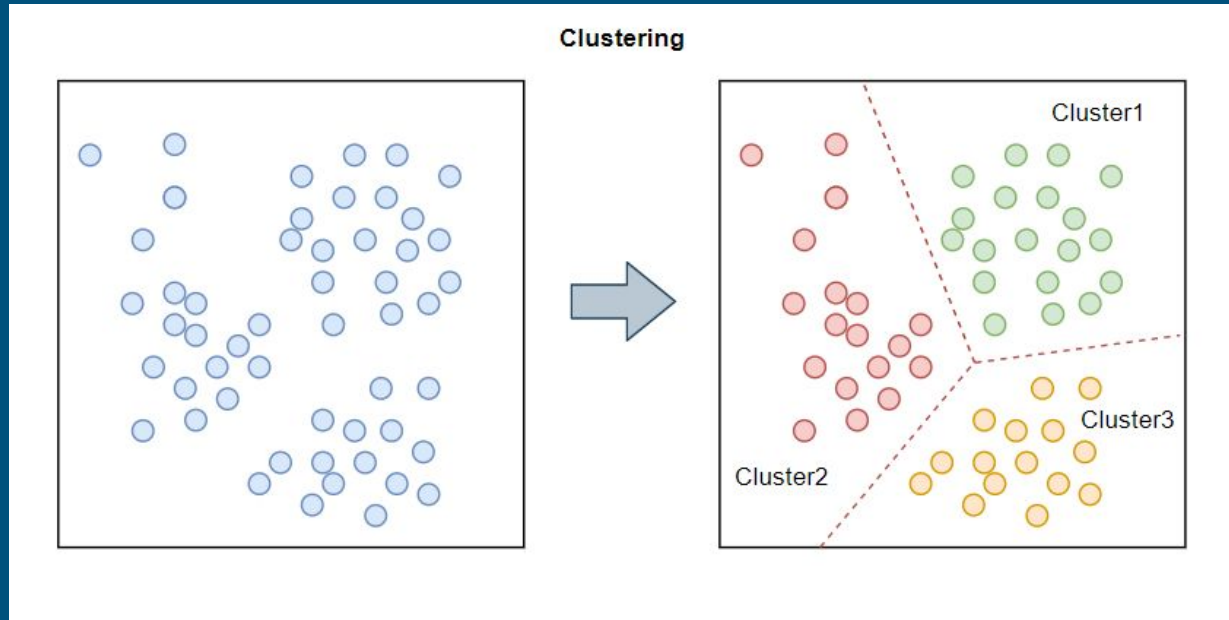


Rótulo Discreto: Classificação



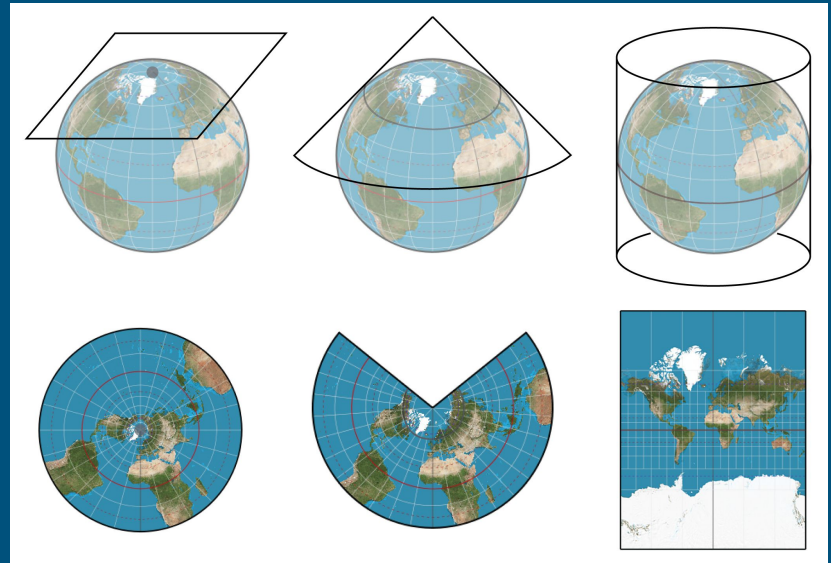
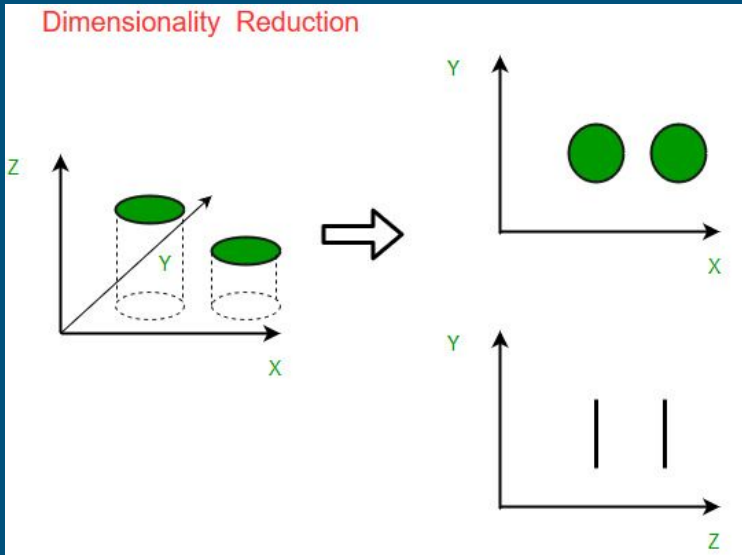
Aprendizado não supervisionado

Agrupamento:

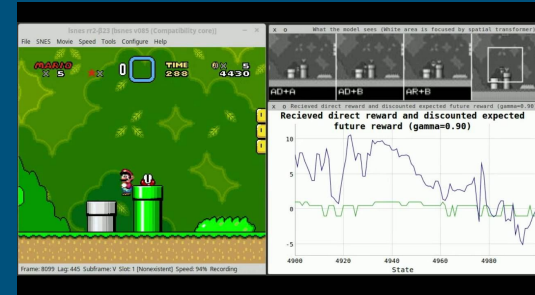
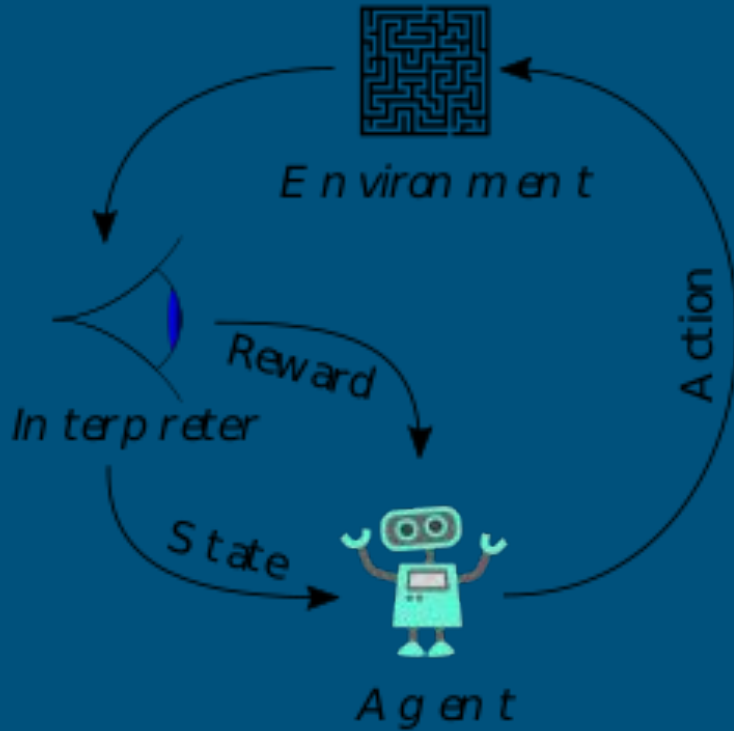


Aprendizado não supervisionado

Redução de dimensionalidade:



Aprendizado reforçado

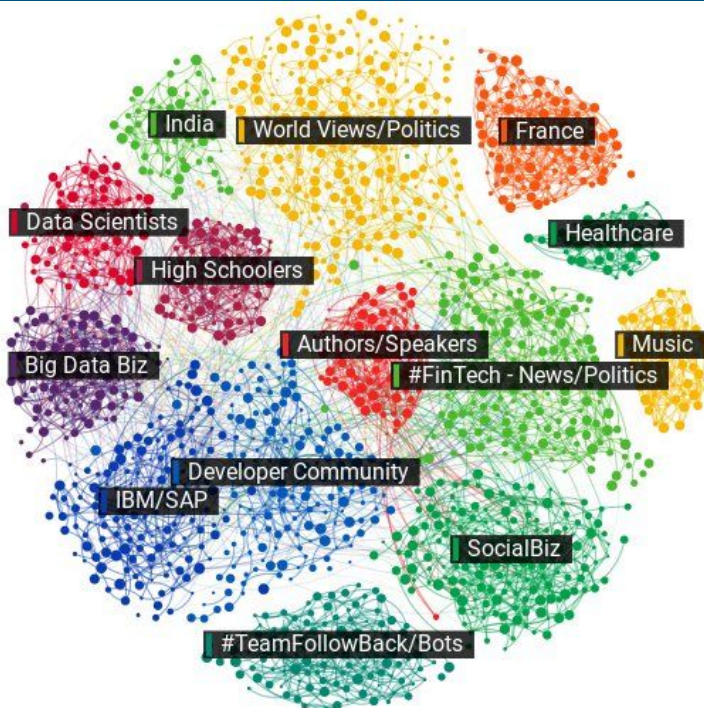




Aplicações

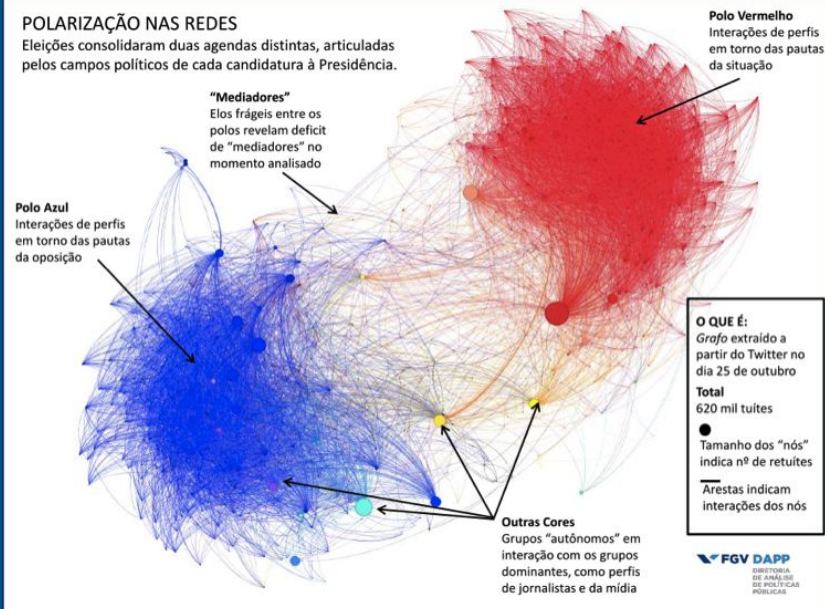


Redes sociais



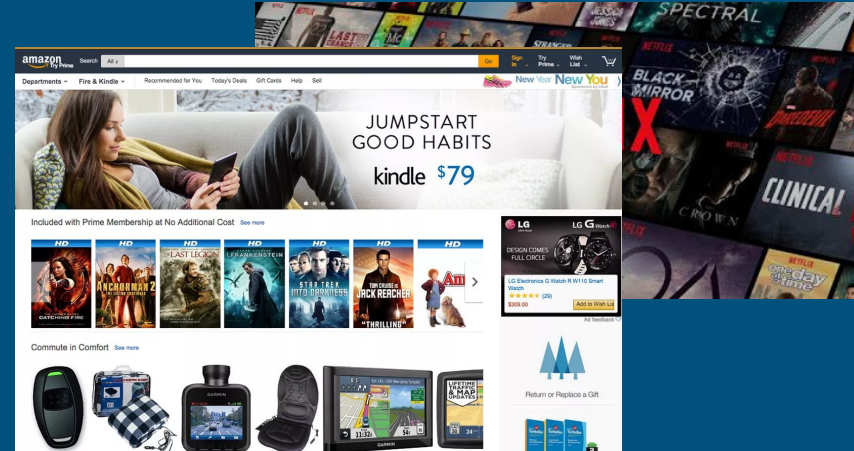
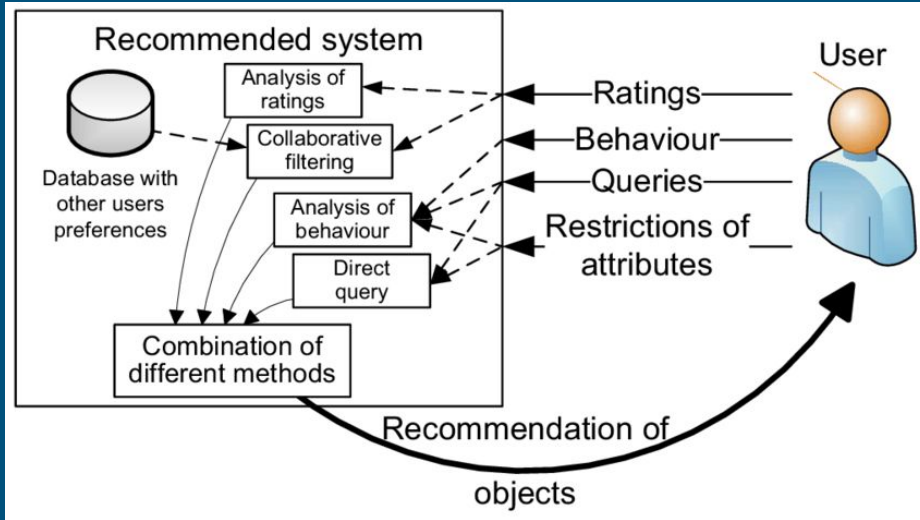
POLARIZAÇÃO NAS REDES

Eleições consolidaram duas agendas distintas, articuladas pelos campos políticos de cada candidatura à Presidência.

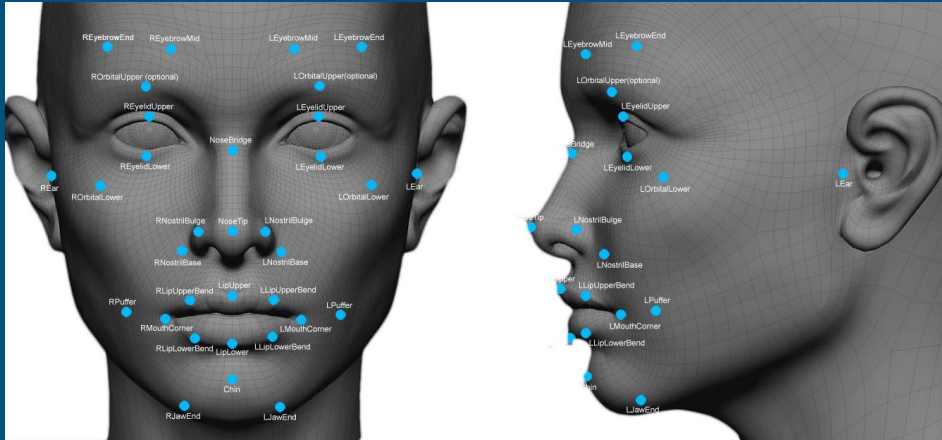


<https://temas.folha.uol.com.br/gps-ideologico/reta-ideologica/a-posicao-ideologica-de-mil-influenciadores-no-twitter.shtml>

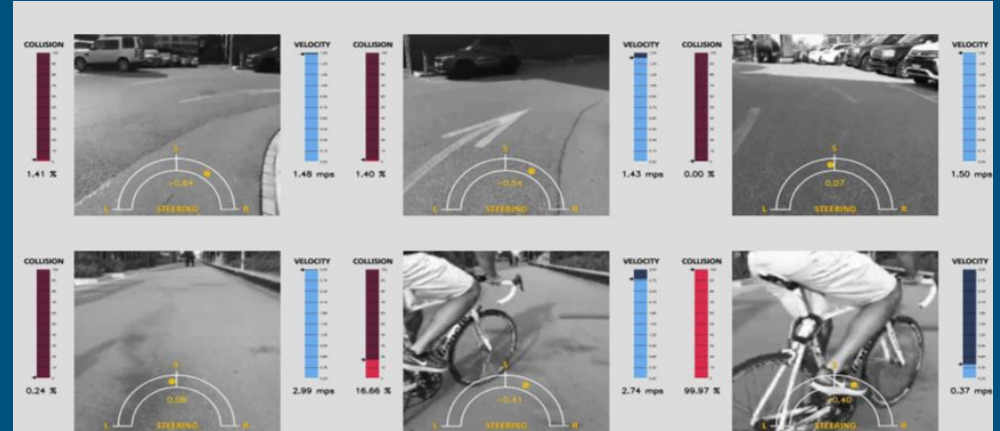
Sistemas de recomendações



Reconhecimento Facial



Piloto automático



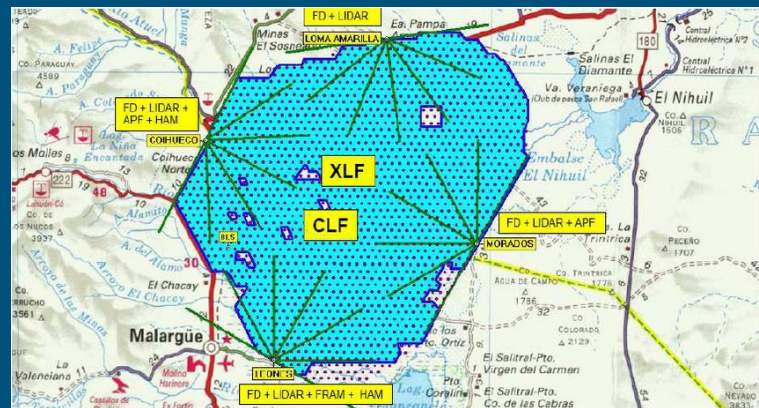
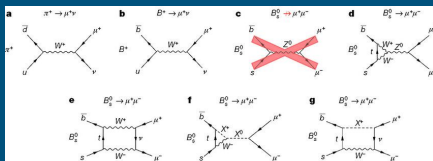
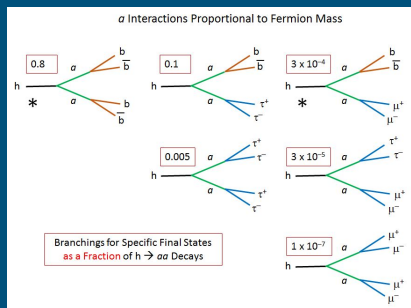
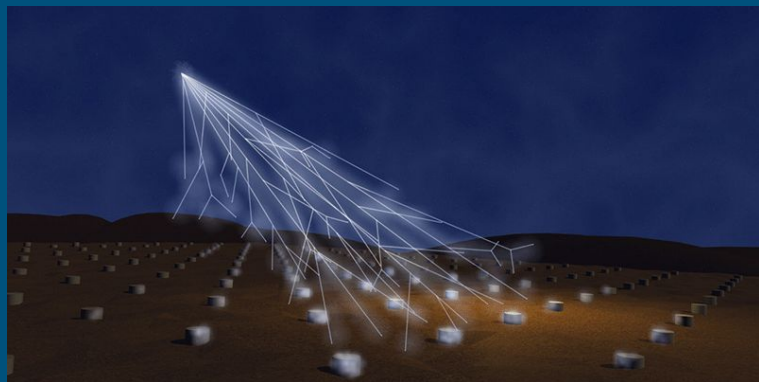
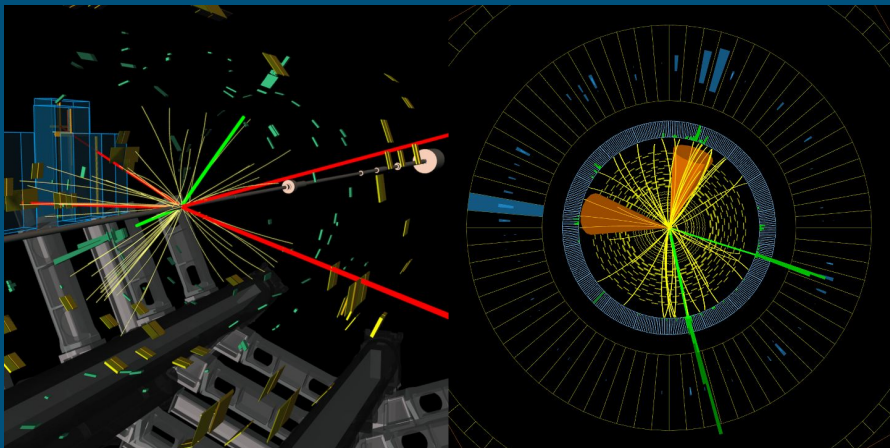
Questões éticas envolvidas

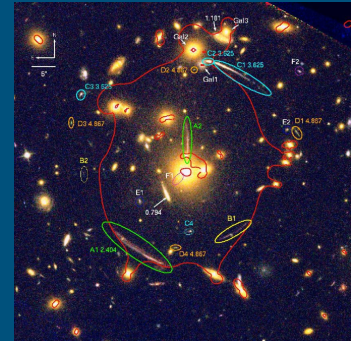
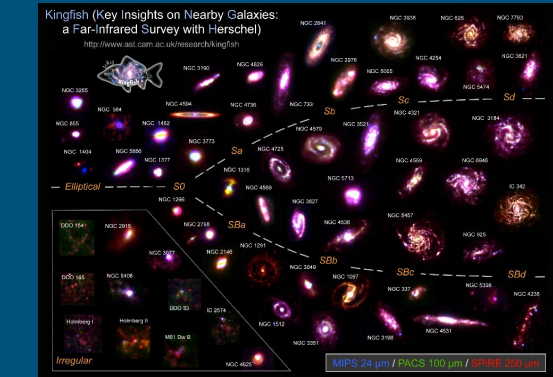


Aplicações na Física

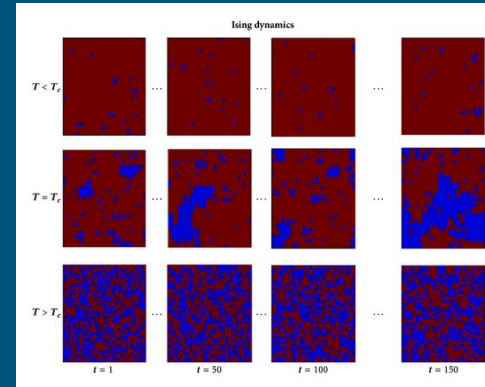
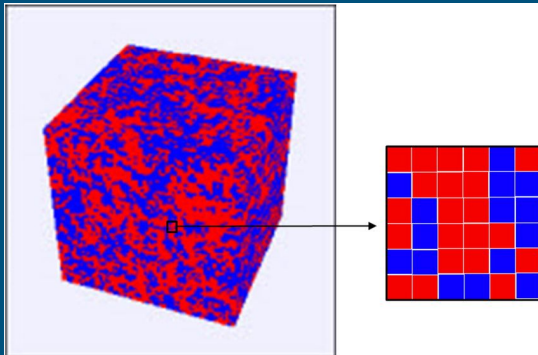
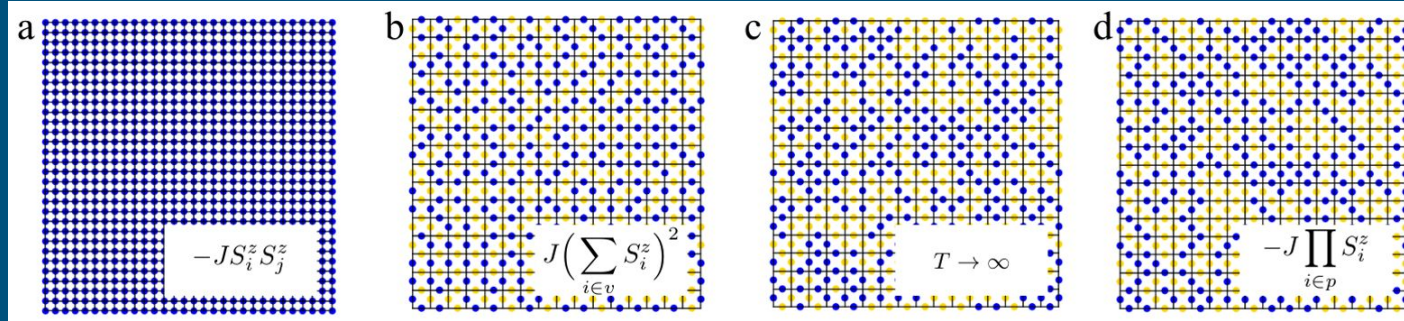


Altas Energias





Estado Sólido e Física Estatística



Análise de sinais

