

Aprendizado de máquina

Introdução

Prof. Jefferson Moraes

Agenda

- Introdução
- Contexto de Aprendizado de máquina
- O que é aprendizado de máquina?
- Por que usar Aprendizado de Máquina?
- Mercado de Aprendizado de Máquina
- Exemplos de Aplicações

Introdução

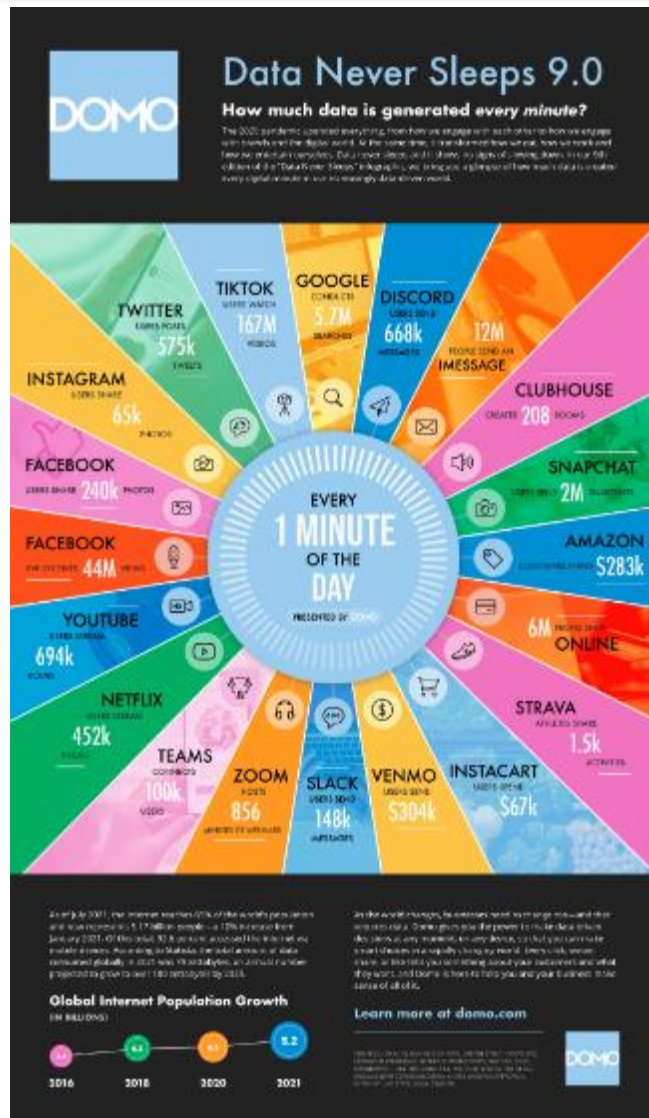
- Pessoas geram dados o tempo todo
 - Aplica para um cartão fidelidade
 - Empresas aérea, supermercado, postos de combustível, etc
 - Faz uma compra com cartão de crédito ou débito
 - Navega pela internet e redes sociais
 - Vai ao médico
- Todos esses dados são armazenados em sistemas computacionais

Explosão de dados

- Vivemos a era do Big Data!
 - Cada ser humano criou cerca de 1.7 MB de dados por segundo em 2020
 - Especialistas indicam que mais de 463 exabytes de dados serão criados a cada dia até 2025
 - 1 exabyte = 1 milhão de terabytes, ou 1 bilhão de gigabytes, ou cerca de 250 milhões de DVDs



Dados nunca dormem



Quanto dados são gerados a cada minuto?

Origem: Domo business management platform

<https://www.domo.com>

2016 a 2021

Análise de dados

- Análise de dados por seres humanos
 - Falta de especialistas
 - Custo elevado
 - Subjetividade
 - Grande volume
- Exemplos de técnicas tradicionais de análise
 - Planilhas eletrônicas
 - Sistemas Gerenciadores de Banco de Dados

Análise de dados

- Técnicas tradicionais de análise de dados respondem perguntas simples do tipo
 - Quantos itens de um produto em particular foram vendidos em um dado dia?
 - Contudo, não conseguem responder
 - Que novo filme poderia recomendar ao assinante?
 - Indicar quando um tecido humano pode apresentar tumor
 - Qual a estrutura terciária de uma nova proteína?
- É necessário uso de técnicas mais sofisticada, capazes de extrair conhecimento de grandes conjuntos de dados
 - Aprendizado de Máquina

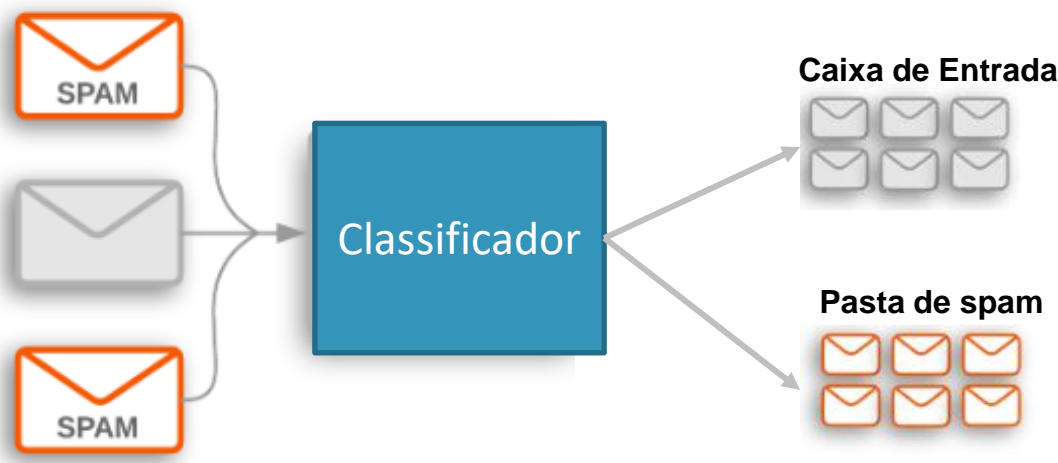
Contexto de AM

- Quando se ouve aprendizado de máquina, logo se imagina



Contexto de AM

- Por outro lado, o aprendizado de máquina não povoa apenas o imaginário, ele já está entre nós
 - Ex: Filtro de spam (década de 1990),



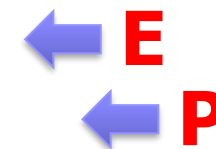
- Centenas de aplicativos, de recomendações a busca por voz, utilizados regularmente
- Mas, o que é aprendizado de máquina?

O que é aprendizado de máquina?

- Algoritmos de AM induzem uma função ou hipótese capaz de resolver o problema a partir de exemplos (instâncias) do problema a ser resolvido – Carvalho et al, 2021
- Aprendizado de máquina é a ciência (e arte) da programação de computadores de modo que eles possam aprender com os dados – Aurélien Géron, 2021
- [Aprendizado de máquina é o] campo de estudo que possibilita aos computadores a habilidade de aprender sem explicitamente programá-los – Arthur Samuel, 1959
- Um programa de computador aprende pela experiência E em relação a algum tipo de tarefa T e alguma medida de desempenho P se o seu desempenho em T , conforme medido por P , melhora com a experiência E – Tom Michell, 1997

Exemplo 1 – Filtro de spam

- Vamos supor que seu programa de email favorito “observa” quais emails você marca ou não marca como spam. Em seguida com base em suas observações (aprendizado) ele consegue uma forma de melhorar o filtro de spam. Defina qual a tarefa T , a experiência E e o desempenho P para o cenário.
 - Classificar emails como spam ou não spam
 - Observar o conjunto de exemplos (treinamento) de spam e não spam
 - O número de emails corretamente classificados como spam e não spam.



Exemplo 2 – Jogar Damas

- Vamos supor o problema de aprender a jogar damas. Identifique a tarefa T, a experiência E e o desempenho P para este problema
 - Porcentagem de vitórias contra oponentes
 - Jogar damas
 - Praticar jogando

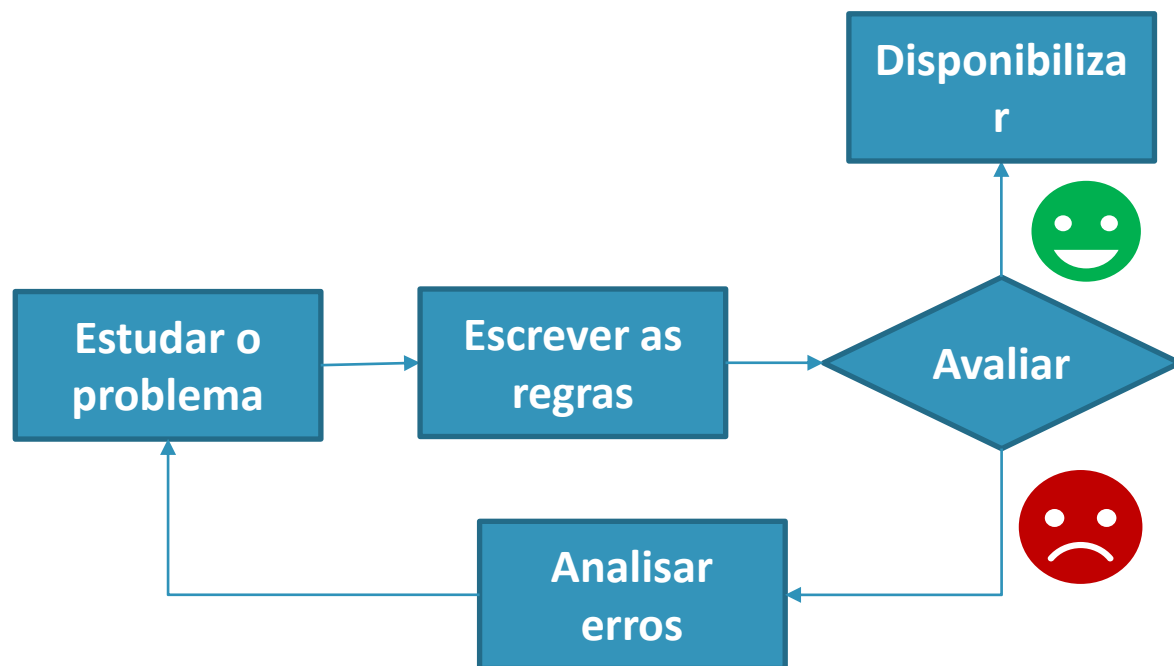
← T

← E

← P

Por que usar AM?

- Como seria o problema do spam usando a abordagem **tradicional de programação**

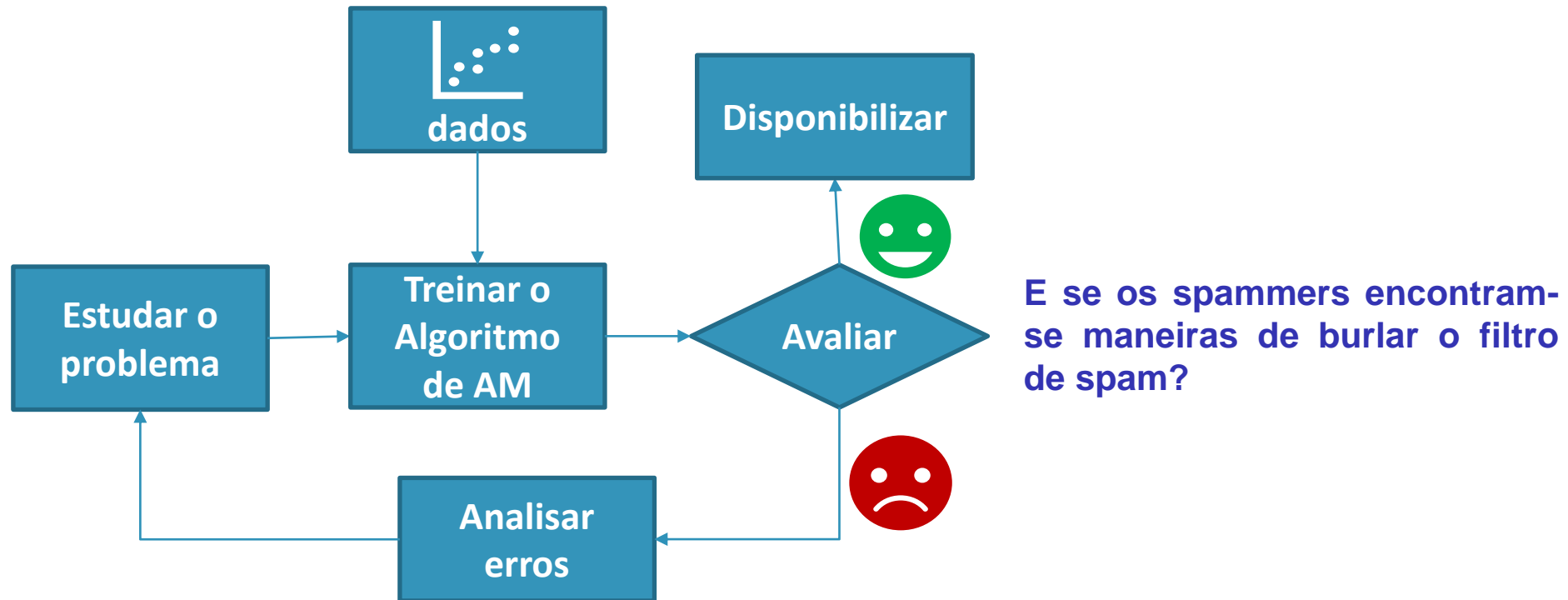


1. Identificar as características de spam (frases: “Para você”, “Cartão de crédito”, etc).
2. Escrever um algoritmo de detecção para cada um dos padrões observados. Uma vez detectados, seriam classificados como spam.
3. Testar o programa e repetir os passos 1 e 2 até que fique bom o suficiente para ser disponibilizado.

- Obs:** Se o problema for difícil, o programa ficará com uma lista extensa de regras, tornando a manutenção uma tarefa nada fácil

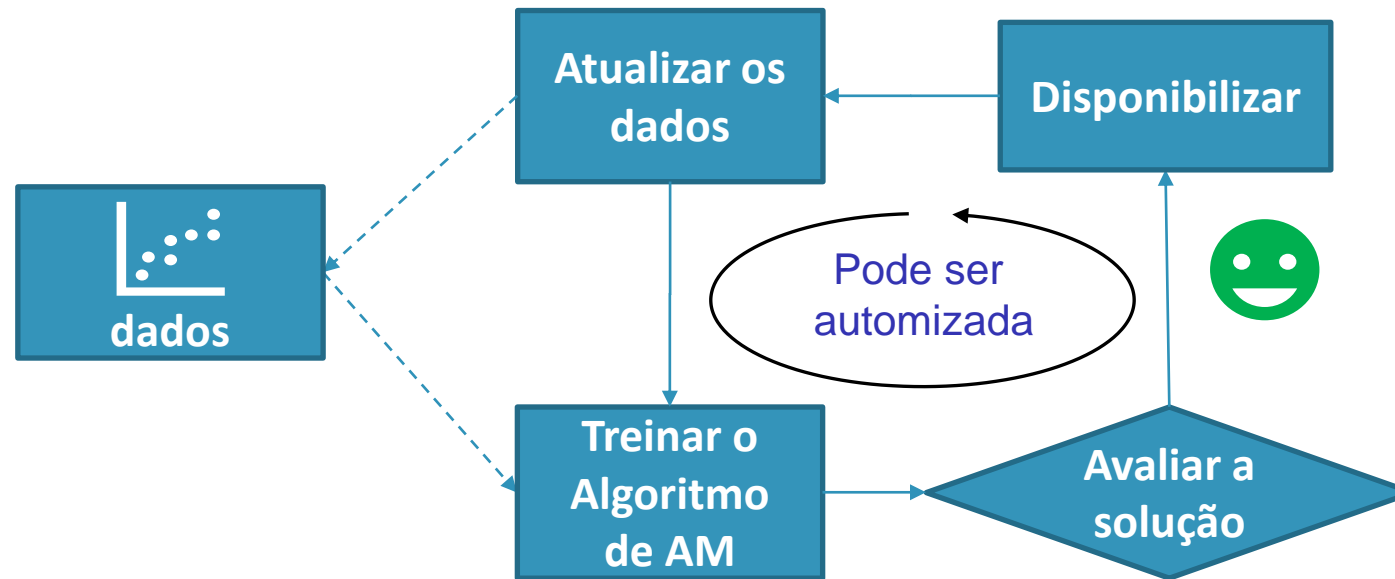
Por que usar AM?

- Filtro de spam baseado em AM aprende automaticamente quais palavras e frases são bons indicadores de spam



Por que usar AM?

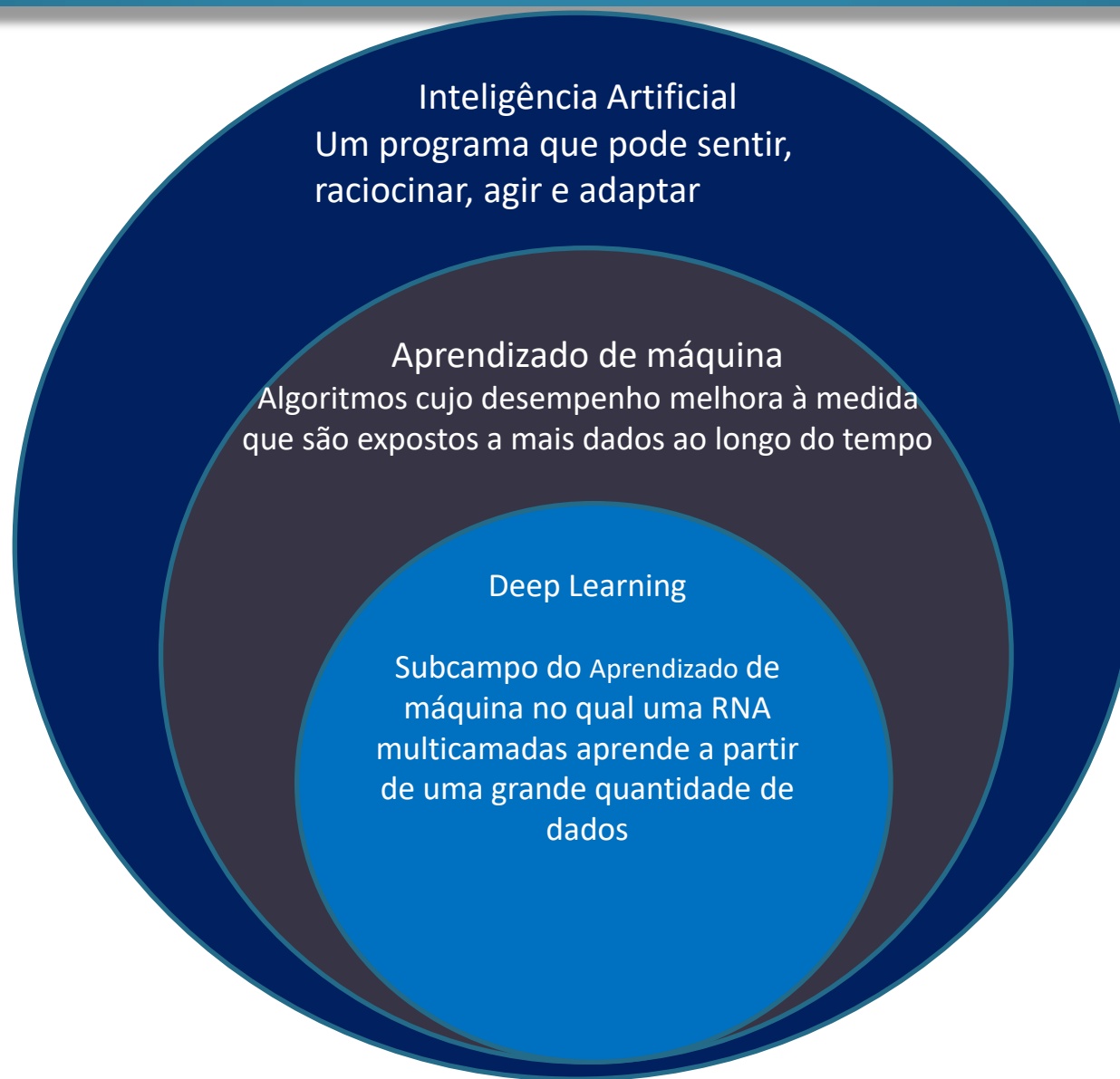
- Modelo que se adapta automaticamente à mudança



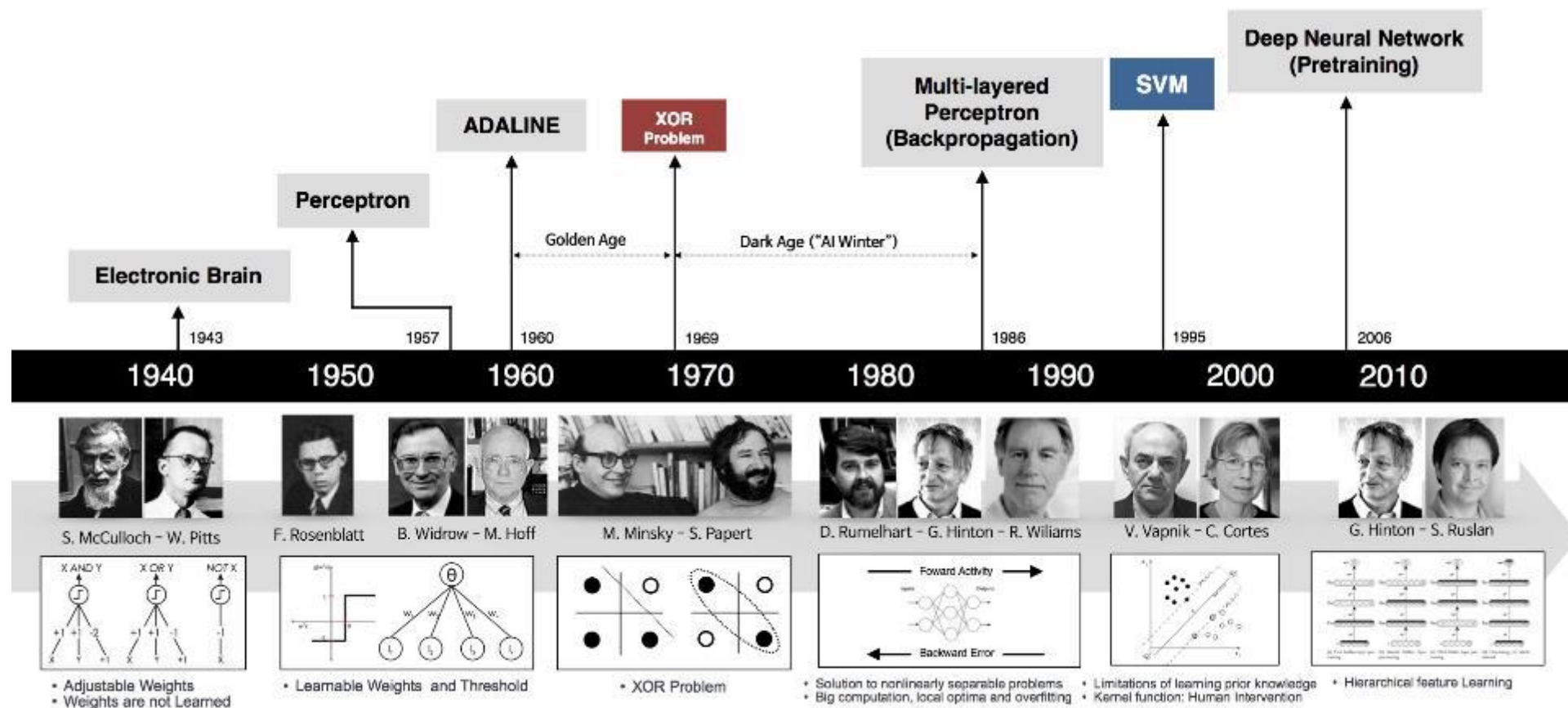
Por que usar AM?

- Resumindo, o AM é ótimo para
 - Problemas que exigem soluções com muitos ajustes finos ou extensas listas de regras
 - Problemas complexos para os quais não existe uma boa solução utilizando a abordagem tradicional
 - Adaptabilidade de ambientes: um sistema de AM pode se adaptar a novos dados
 - Entendimento de problemas complexos e grande quantidade de dados

AM subcampo da IA



AM time line



Motivações

- Barateamento e popularização dos dispositivos móveis e dos mais diversos tipos de sensores para coleta de dados
- Aumento do poder de processamento e armazenamento dos computadores
- Otimização de recursos em um mercado global cada vez mais competitivo
- Amadurecimento, melhoria e disseminação de técnicas de aprendizado de máquina

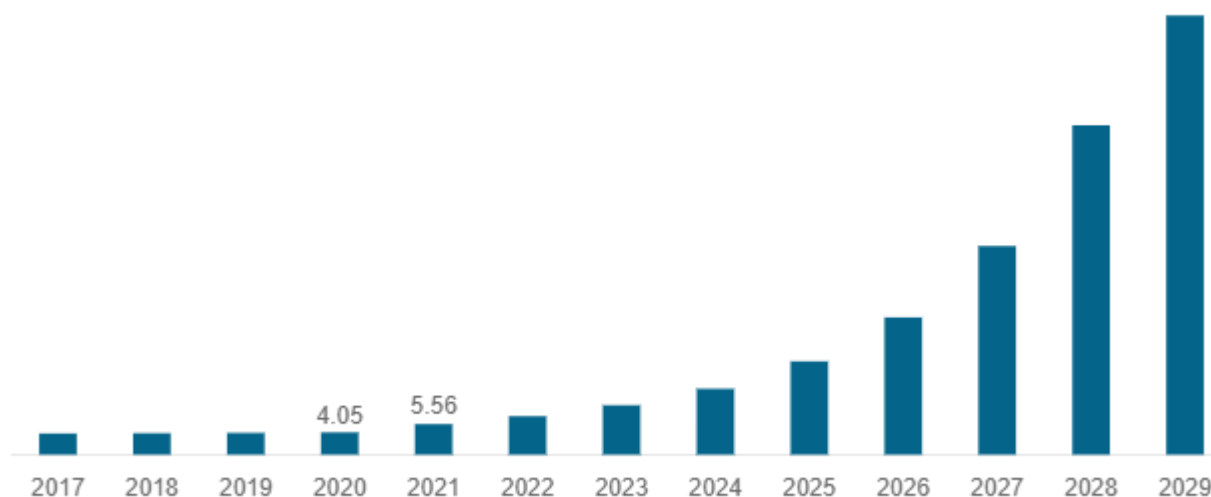


Mercado de AM

- O mercado de AM está experimentando um **crescimento exponencial** em todo o mundo
- Segundo a Fortune é projetado para crescer de US\$ 15,50 bilhões em 2021 para US\$ 152,24 bilhões em 2028 - uma taxa de crescimento anual de 38,6% no período 2021-2028



North America Machine Learning (ML) Market Size, 2018-2029 (USD billion)



[www.fortunebusinessinsights.com](https://www.fortunebusinessinsights.com/machine-learning-market-102226)

<https://www.fortunebusinessinsights.com/machine-learning-market-102226>

Mercado de AM (Brasil)

- De acordo com dados da IDC Brasil as empresas estão se dando conta da importância de se ter insights sobre seu próprio volume de dados
- Em 2022, a IDC Brasil estima que U\$ 504 milhões serão destinados a soluções e serviços relacionados a inteligência artificial e machine learning, um crescimento de 28% em relação a 2021.



4- Uso de dados para impulsionar negócios colocará Analytics, AI/ML e Data Management na pauta prioritária de mais de 47% das empresas. Edge e 5G contribuem para este movimento



Cada vez mais as empresas compreendem a necessidade de se ter insights acionáveis sobre seus próprios dados, especialmente com a ascensão dos canais digitais. Em 2022, a IDC estima que U\$ 2,9 bilhões serão destinados a soluções e serviços relacionados a Big Data & Analytics no Brasil, um aumento de 10,8% em relação ao ano passado. Já em Inteligência Artificial e Machine Learning, a expectativa da IDC é de um crescimento de 28% no mesmo período, chegando a U\$ 504 milhões.

“Mesmo havendo uma evolução clara, muitas empresas ainda carecem de uma cultura de dados estabelecida e difundida nas áreas de negócios, o que é necessário para que se possam estabelecer métricas e modelos de tomada de decisão ágeis, gerando maior competitividade. Neste contexto, o avanço do edge computing, estimulado pelo 5G, deve ocupar papel de destaque”, conclui Ramos.

<https://www.idc.com/getdoc.jsp?containerId=prLA49041022>

Mercado de AM

- De acordo com o Fórum Econômico Mundial, a inteligência artificial e o machine learning impulsionarão a criação de mais de 97 milhões de novos empregos até 2025

[Find jobs](#) [Company reviews](#) [Find salaries](#)

Machine learning engineer: breaking into the best job of 2019

As artificial intelligence (AI) and machine learning (ML) reshape how we live and the things we use every day like our cars, homes and mobile devices, the engineers who work with these technologies have quickly grown in demand. With an average base salary of \$146,085 and a rocketing 344% growth in job postings, Indeed's [study of jobs experiencing the fastest growth](#) and offering the highest pay for 2019 reveals that machine learning has moved up to the top spot from number four last year.

<https://www.indeed.com/career/machine-learning-engineer/news>

Mercado de AM

<https://insights.stackoverflow.com/survey/2021#salary-comp-total>



Exemplos de Aplicações

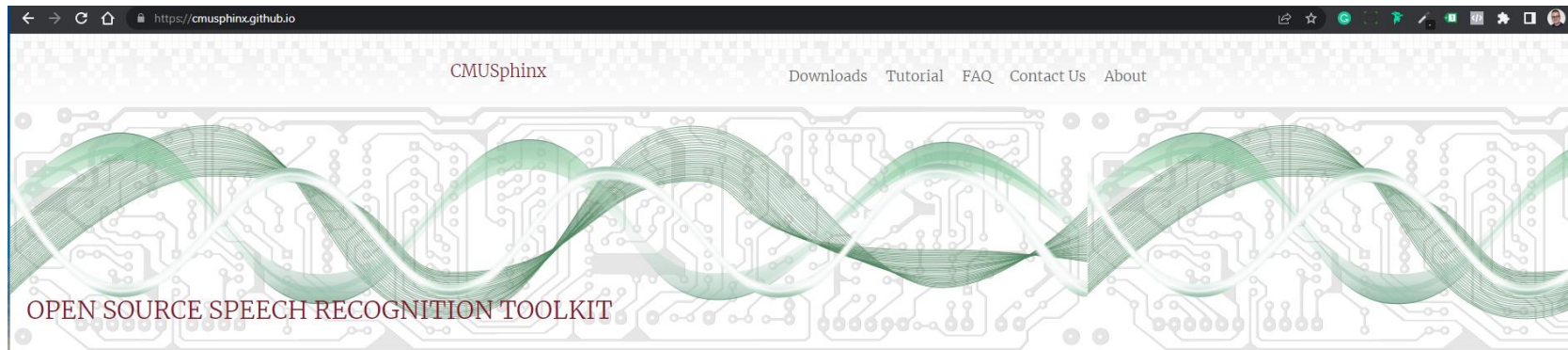


Exemplos de Aplicações

- Programas baseados em AM têm sido bem sucedidos para
 - Análise de redes sociais
 - Análise de dados biológicos
 - Detecção de fraudes
 - Diagnóstico médico
 - Reconhecimento facial e de voz
 - Sistemas de recomendação

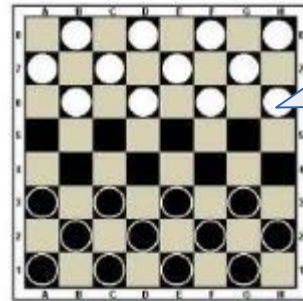
Exemplos de Aplicações

- Aprender a reconhecer palavras faladas
 - SPHINX



Exemplos de Aplicações

- Aprender a **jogar** damas
 - Tarefa T: jogar damas
 - Medida de desempenho P: porcentagem de jogos vencidos contra adversários ou contra si próprio
 - Experiência E: praticar jogando



Arthur Lee Samuel
(criou o termo "*machine learning*" em 1959):
programa de jogar damas,
primeiro programa com auto-aprendizado.
Jogou milhares de vezes contra si mesmo
Chegou a nível amador

Exemplo de Aplicações



Exemplos de Aplicações

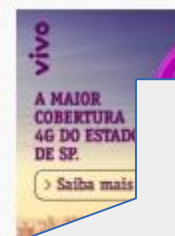
Spam

Google says its machine learning tech now blocks 99.9% of Gmail spam and phishing messages

Posted May 31, 2017 by Frederic Lardinois (@frednic)



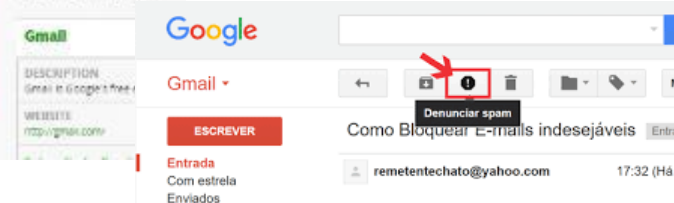
Google today said that its machine learning models can now detect spam and phishing messages with 99.9 percent accuracy.



50-70 percent of messages that Gmail receives are spam

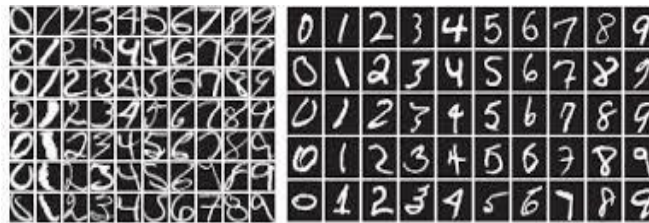


Crunchbase



Exemplos de Aplicação

- **Reconhecer** escrita manual
 - **Tarefa T**: reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P**: ?
 - **Experiência E**: ?

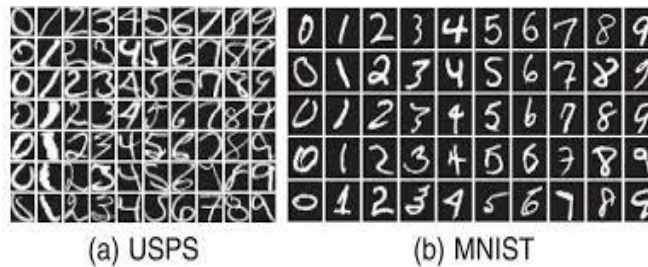


(a) USPS

(b) MNIST

Exemplos de Aplicações

- **Reconhecer** escrita manual
 - **Tarefa T**: reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P**: porcentagem de dígitos corretamente identificados
 - **Experiência E**: imagens de dígitos manuscritos por diferentes pessoas



USPS database: dados numéricos obtidos da digitalização de dígitos manuscritos de envelopes pelo Serviço Postal dos EUA

MNIST database:
Imagens de dígitos manuscritos, coletados entre os funcionários do Censo e alunos do ensino médio.

Exemplos de Aplicações

Deep learning e gatos

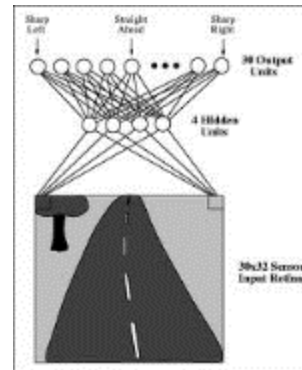
“When computer scientists at Google’s mysterious X lab built a neural network of 16,000 computer processors with one billion connections and let it browse YouTube, it did what many web users might do — it began to look for cats”

WIRED STAFF SCIENCE 06/26/12 11:15 AM
**GOOGLE’S ARTIFICIAL BRAIN
LEARNS TO FIND CAT VIDEOS**



Exemplo de Aplicações

- Carro autônomo (aprender a **dirigir**)
 - Tarefa T: dirigir em uma rodovia pública usando diversos sensores
 - Medida de desempenho P: ?
 - Experiência E: ?

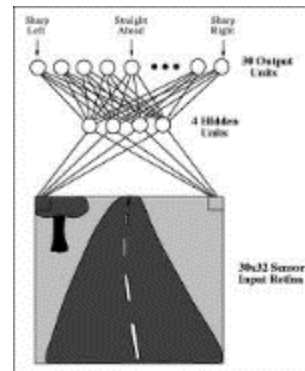


*ALVINN, an
autonomous land
vehicle in a neural
network*

Exemplos de Aplicações

carro autônomo (aprender a **dirigir**)

- **Tarefa T:** dirigir em uma rodovia pública usando diversos sensores
- **Medida de desempenho P:** distância média percorrida antes de um erro
- **Experiência E:** sequência de imagens e comandos de direção registrados observando um motorista humano



Exemplos de Aplicação

Self-Driving Car



Self-Driving Trucks

Tractor-trailers without a human at the wheel will soon barrel onto highways near you. What will this mean for the nation's 1.7 million truck drivers?

Availability: 5 to 10 years



Exemplos de Aplicações

- **Detectar** bons clientes
 - **Tarefa T**: classificar potenciais clientes como bons ou maus pagadores
 - **Medida de desempenho P**: ?
 - **Experiência E**: ?



Exemplos de Aplicações

- **Detectar** bons clientes
 - **Tarefa T**: classificar potenciais clientes como bons ou maus pagadores
 - **Medida de desempenho P**: porcentagem de clientes classificados corretamente como bons e maus pagadores
 - **Experiência E**: uma base de dados histórica com dados de clientes já conhecidos



Exemplos de Aplicações

- Detecção de fraudes



Reduza fraudes e aumente vendas

A Ravelin usa a aprendizagem de máquina para identificar precisamente contas e ordens ruins em seus dados. É a maneira mais eficaz de permitir mais clientes genuínos comprarem a partir de sua loja on-line ou app, eliminando estornos. Venda com confiança



Outras aplicações clássicas AM

- Número crescente de aplicações
 - **Finanças**: análise de risco, detecção de fraudes, gerenciamento de carteiras
 - **Internet**: algoritmos de busca, marketing na web
 - **Ciência e Medicina**: descoberta de padrões, diagnóstico de pacientes, análise de dados do genoma
 - **Indústrias**: previsão de falhas, diagnóstico de produtos
 - **Marketing**: segmentação de mercado, recomendação de produtos
 - **Telecomunicações**: processamento de alarmes e sensores

Algumas Ferramentas

Rapid Miner

- <http://rapidminer.com>



Weka

- <http://www.cs.waikato.ac.nz/ml/weka>



Keel

- <http://www.keel.es/>



Python – Scikit-learn:

- <http://scikit-learn.org/>



R Project

- <http://www.r-project.org>



Repositório de dados

- Machine Learning Data Repository UC Irvine

<http://www.ics.uci.edu/~mlern/ML/Repository.html>



- Kaggle: competições práticas promovidas por empresas

<https://www.kaggle.com/competitions>

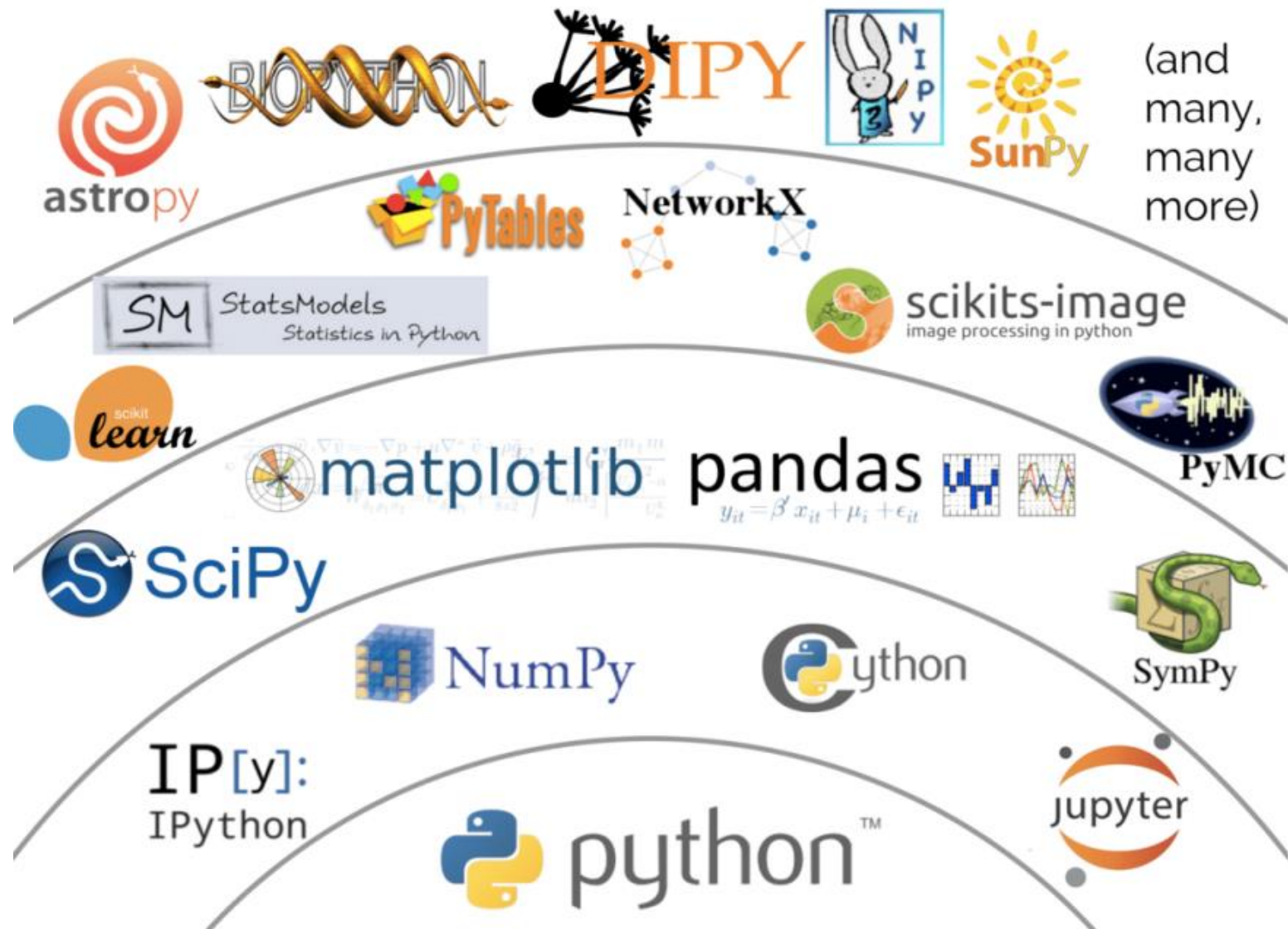


- OpenML

<https://www.openml.org/>



Ecossistema Python



Referências

- Alguns slides foram baseados em
 - Apresentações de
 - Profa. Dra. Ana Carolina Lorena
 - Prof. Dr. Regis Magalhães
 - Prof Dr André C. P. L. F. Carvalho
 - Livros
 - Inteligência Artificial: uma abordagem de Aprendizado de Máquina, 2021
 - Machine Learning (Tom Mitchel - 1997)
 - Mão à Obra: Aprendizado de Máquina com scikit-learn, Keras e TensorFlow, 2021

Atividade 1

- Faça o resumo do seguinte artigo: “Some studies in Machine Learning Using the Game of Checkers”
- Com base no artigo “The Discipline of Machine Learning” e responda as seguintes perguntas:
 - É possível que dados não rotulados sejam úteis para aprendizado supervisionado?
 - Como se pode transferir aquilo que é aprendido em uma tarefa para outra tarefa relacionada?
 - Qual é a relação existente entre diferentes algoritmos de aprendizado e qual deve ser usado?
 - Pode-se criar máquinas que nunca param de aprender?
- Data de entrega: Próxima aula