Quem sou eu ... filha, irmã, tia, stepmother, mãe de planta...aspirante a peregrina

- Graduação em Ciência da Computação pela Universidade Católica de Goiás
- Especialização em Redes de Computadores pela Universidade Salgado de Oliveira
- MBA em Gerenciamento de Projetos pela Fundação Getúlio Vargas
- Mestrado em Ciência da Computação pela Universidade Federal de Goiás
- Doutorado em Ciência da Computação pela Universidade Federal de Goiás
- MBA em **Tecnologia para Negócios: Al, Data Science e Big Data** pela Pontifícia Universidade Católica do Rio Grande do Sul
- Professora/coordenadora dos cursos de graduação: Ciência da Computação, Engenharia da Computação, Sistemas de Informação, Análise e Desenvolvimento de Sistemas, Data Science, Defesa Cibernética e Gestão de Tecnologia (UNIALFA, ESTÁCIO, UNIP, PUCGOIÁS, FASAM, UNIANHANGUERA, IPOG)
- **Professora de cursos de pós graduação**: Engenharia de Software, Big Data e Machine Learning (UniAnhaguera, Faculdade Sul-Americana)
- Atualmente Analista Judiciário Ciência de Dados no Tribunal de Justiça do Distrito Federal e Territórios, lotada na SUDES

UNIDADE 1: Inteligência Artificial

- 1.1. Fundamentos do Aprendizado de Máguina
- 1.2. Tipos de Tarefas de Aprendizado
- 1.2.1. Supervisionado
- 1.2.2. Não supervisionado
- 1.2.3. Semi supervisionado
- 1.2.4. Aprendizado por reforço
- 1.3. Modelagem preditiva

Prática (TOY): Classificação de casos de Alzheimer (dados fictícios)

UNIDADE 3: Regressão Linear

- 3.1 Previsões simples (Regressão linear)
- 3.2 Previsões complexas (Regressão linear múltipla)
- 3.3 Algoritmos de Classificação
- 3.3.1 Métodos simbólicos: Árvore de Decisão
- 3.3.2 Métodos ensemble: Random Forest

Prática: Previsão da carga de trabalho dos juízes (dados fictícios)

UNIDADE 5: Redes Neurais

- 5.1 Estrutura básica de uma rede neural
- 5.2 Arquitetura de redes neurais
- 5.2.1 Perceptron
- 5.2.2 Multilaver Perceptron
- 5.3 Treinamento de redes neurais
- 5.3.1 Algoritmos de retropropagação
- 5.3.2 Otimização de pesos
- 5.3.3 Função de ativação
- 5.4 Arquiteturas de Redes Neurais Profundas

Prática: Predição do tempo de atravessamento dos processos (dados fictícios)

UNIDADE 2: Tratamento dos dados

- 2.1. Análise exploratória de dados (EDA)
- 2.2. Limpeza de dados
- 2.3. Transformação de dados
- 2.3.1. Divisão dos dados
- 2.3.2. Balanceamento
- 2.3.3. Escalonamento
- 2.3.4. Codificação e criação de features
- 2.4. Redução de dimensionalidade

Prática: Planejamento de sucessão cargos de liderança (dados fictícios)

UNIDADE 4: Otimização e desempenho

- 4.1 Otimização de hiperparâmetros e regularização
- 4.1.1 Técnicas de avaliação de modelos
- 4.1.1.1 Overfitting e underfitting
- 4.1.1.2 Interpretabilidade dos modelos
- 4.1.1.3 Questões éticas e de viés
- 4.1.2 Validação cruzada
- 4.1.3 GridSearchCV
- 4.1.4 Avaliação de métricas

Prática: Previsão da carga de trabalho dos juízes (refinamento)

UNIDADE 6: Large Language Models

- 6.1 Redes Neurais Generativas (GAN)
- 6.2 Conceitos de Processamento de Linguagem Natural
- 6.3 Principais modelos LLM
- 6.3.1 Open Al
- 6.3.2 Llama

Prática: Chatbot Geral.

Reconhecimento de Entidade

Nomeada, Fine-Tunning do Llama3

UNIDADE 7: Treinando Modelos LLMs

- 7.1 Retrieval Augmented Generation
- 7.2 Framework LangChain
- 7.3 Llama 3

Prática: Chatbot na base de informações da ouvidoria do TJDFT (dados reais)





Aprendizado Supervisionado e Não Supervisonado

 Prática: Classificação de casos de Alzheimer Análise exploratória de dados (EDA), Limpeza e Transformação de dados

 Prática: Planejamento de sucessão (Classificação) Regressão Linear, Árvores de Decisão, Regularização, Métricas de desempenho

 Prática: Predição de Cargas de trabalho para juízes



Validação cruzada, otimização de hiperparâmetros e AutoML



Redes Neurais

 Prática: Predição do tempo de atravessamento de ações na justiça

Principais pontos...

Retrieval Augmented Generation (RAG)

- Prática 1: Chatbot Ouvidoria (Nvidia + Qdrant + Llama3)
- Prática 2: Chatbot Ouvidoria (AnythingLLM + LMStudio + Llama3)



Large Language Models

- · Prática 1: Chatbot geral Ollama+Llama3
- Prática2: Reconhecimento de entidade nomeada Groq+Llama3
- Prática 3: Reconhecimento de entidade nomeada Spacy
- Prática 4: Fine-tuning Llama3

Material e Ambiente

Ambiente de compartilhamento de códigos, datasets, slides e referências adicionais usados nas aulas:

- Repositório no GitHub (https://github.com/joelmaf/course.ai). À medida que as aulas forem avançando esse repositório será atualizado
- Para entrega das atividades usar o AVA
- Ambiente de execução
 - Instalar na sua máquina: Anaconda Navigator
 - Usar na nuvem: Google Colab

Agenda

UNIDADE 1: Inteligência Artificial

- Fundamentos do Aprendizado de Máquina
- Tipos de Tarefas de Aprendizado
 - Supervisionado
 - Não supervisionado
 - Semi supervisionado
 - Aprendizado por reforço
- Modelagem preditiva

Ciência (análise) de Dados (na opinião de especialistas)

- Ciência de dados é um processo (Shingau Manjengwa, CEO, Fireside Analytics)
- Validação de hipóteses e modelos através de dados (Rafael Silva, Data Scientist)
- Descobrir insights escondidos nos dados Diana Diaz, Data Scientist)
- Tradução de dados em histórias (Luis O. Martins, Data Scientist)
- É o estudo dos dados (Stephen Sherman, Data Scientist)
- Envolve ciência e dados (Murtaza Haider, PhD)

Grande quantidade de dados



Capacidade computacional



Análise e novos conhecimentos

Ciência (análise) de Dados (base)

• Em geral, foco em um problema específico

Clarificar um problema que a empresa que resolver

- Quais dados precisamos para resolver o problema?
- Quais as fontes (estruturado ou não estruturado) de dados?
- Quais modelos serão utilizados para explorar (padrões e/ou anomalias) os dados?



Habilidades Essenciais

- Curiosidade
 - O que fazer com os dados
- Habilidade de julgamento
 - Por onde começar
- Habilidade de argumentação
 - Aprender com os dados
- Habilidade de contar histórias
 - Storytelling

You learn from data!!!!

Como Resolver a Situação?

Classificação de casos de Alzheimer

ID	M/F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay
OAS1_0001_MR1	F	R	74	2.0	3.0	29.0	0.0	1344	0.743	1.306	NaN
OAS1_0002_MR1	F	R	55	4.0	1.0	29.0	0.0	1147	0.810	1.531	NaN
OAS1_0003_MR1	F	R	73	4.0	3.0	27.0	0.5	1454	0.708	1.207	NaN
OAS1_0004_MR1	М	R	28	NaN	NaN	NaN	NaN	1588	0.803	1.105	NaN
OAS1_0005_MR1	M	R	18	NaN	NaN	NaN	NaN	1737	0.848	1.010	NaN

A escala clínica usada para avaliar o estágio de gravidade da demência em pacientes: demência muito leve, demência leve, demência moderada, além de sem demência

['CDR'] == 0.0] = "SD" ['CDR'] == 0.5] = "DML" ['CDR'] == 1.0] = "DL" ['CDR'] == 2.0] = "DM"







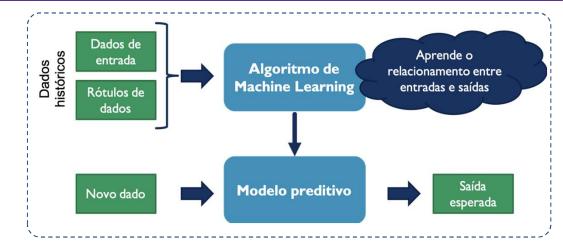


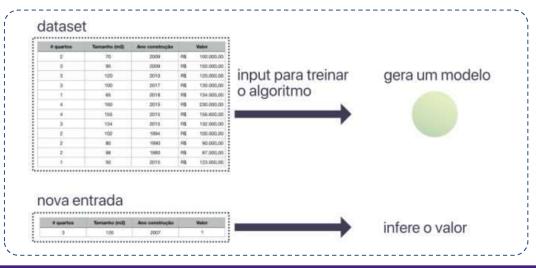


- ☐ ID (identificação)
- ☐ Sexo (M/F)
- ☐ Mão dominante (Hand)
- ☐ Idade (Age)
- Nível de educação (Edu)
- ☐ Status socioeconômico (SES)
- ☐ Pontuação no Mini-Mental State Examination(MMSE) que avalia a função cognitiva
- Classificação do Clinical Dementia Rating (CDR) que indica o estágio de gravidade da demência
- ☐ Volume total do cérebro e do crânio (eTIV)
- □ Volume Cerebral Total Normalizado (nWBV) que mede o volume cerebral relativo ao tamanho total do cérebro
- □ Valor único do tamanho relativo do cérebro em relação a um atlas padrão (ASF)
- □ Atraso (Delay), medido em meses, entre a aquisição das imagens cerebrais e a avaliação clínica ou cognitiva dos participantes

Aprendizado de Máquina

- Compreensão dos padrões presentes em um conjunto de dados.
- Algoritmos s\(\tilde{a}\) alimentados com grandes quantidades de dados.
- Técnicas de ML são capazes de:
 - treinar sua inteligência por meio de dados de entrada
 - criar modelos que conseguem tomar decisões baseadas nos resultados obtidos por meio do treinamento.

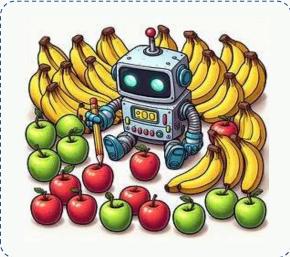




Supervisionado e Não Supervisionado

Os dados contêm a resposta desejada (rotulados)



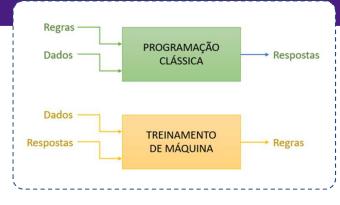


O algoritmo não será supervisionado, não há uma resposta prévia.

Aprendizado Supervisionado

DADOS DE ENTRADA (TREINO DO MODELO)

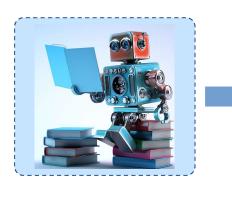
Profissão	Salário	Tipo de Dívida	Rótulo (Pagou?)
Médico	15000	Consignado	Sim
Estudante	1000	Empréstimo X	Não
Apostentado	1200	Consignado	Sim
Economista	5000	Empréstimo X	Sim



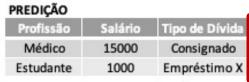
Sim

Não





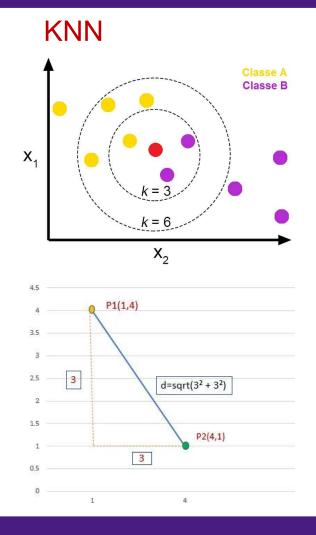


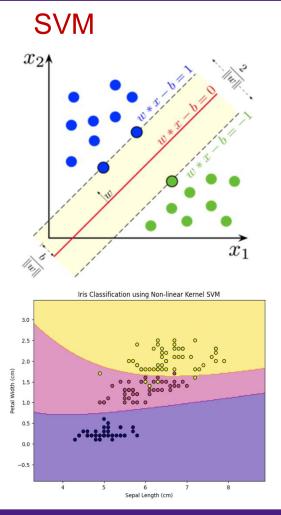


NOVOS DADOS

Profissão	Salário	Tipo de Dívida
Médico	15000	Consignado
Estudante	1000	Empréstimo X

Aprendizado Supervisionado







E agora...como Resolver a Situação?

Classificação de casos de Alzheimer

ID	M/F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay
OAS1_0001_MR1	F	R	74	2.0	3.0	29.0	0.0	1344	0.743	1.306	NaN
OAS1_0002_MR1	F	R	55	4.0	1.0	29.0	0.0	1147	0.810	1.531	NaN
OAS1_0003_MR1	F	R	73	4.0	3.0	27.0	0.5	1454	0.708	1.207	NaN
OAS1_0004_MR1	M	R	28	NaN	NaN	NaN	NaN	1588	0.803	1.105	NaN
OAS1_0005_MR1	M	R	18	NaN	NaN	NaN	NaN	1737	0.848	1.010	NaN

A escala clínica usada para avaliar o estágio de gravidade da demência em pacientes: demência muito leve, demência leve, demência moderada, além de sem demência

['CDR'] == 0.0] = "SD" ['CDR'] == 0.5] = "DML" ['CDR'] == 1.0] = "DL" ['CDR'] == 2.0] = "DM"











- ☐ ID (identificação)
- ☐ Sexo (M/F)
- ☐ Mão dominante (Hand)
- ☐ Idade (Age)
- Nível de educação (Edu)
- ☐ Status socioeconômico (SES)
- ☐ Pontuação no Mini-Mental State Examination(MMSE) que avalia a função cognitiva
- Classificação do Clinical Dementia Rating (CDR) que indica o estágio de gravidade da demência
- ☐ Volume total do cérebro e do crânio (eTIV)
- Volume Cerebral Total Normalizado (nWBV) que mede o volume cerebral relativo ao tamanho total do cérebro
- □ Valor único do tamanho relativo do cérebro em relação a um atlas padrão (ASF)
- □ Atraso (Delay), medido em meses, entre a aquisição das imagens cerebrais e a avaliação clínica ou cognitiva dos participantes



Ciência de dados é um processo

Shingau Manjengwa, CEO, Fireside Analytics

É o estudo dos dados

Stephen Sherman, Data Scientist

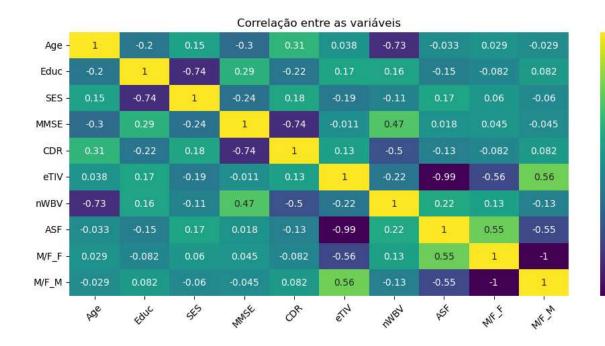
Passo 1: Limpeza e Transformação

- Delay, ID e Hand: colunas descartadas (dados faltosos sem valor estatístico), só tem destros.
- 2. Removidas as linhas que apresentem:
 - Dado faltante
 - Demência mais avançada 2.0, só tem dois casos (descartado)
- M/F" transformada em quantitativa utilizando o método de codificação One Hot-Encoder
- 2. Padronização dos dados

	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	M/F_F	M/F_M	CDR_
0	74	2.0	3.0	29.0	0.0	1344	0.743	1.306	1.0	0.0	SD
1	55	4.0	1.0	29.0	0.0	1147	0.810	1.531	1.0	0.0	SD
2	73	4.0	3.0	27.0	0.5	1454	0.708	1.207	1.0	0.0	DML
3	74	5.0	2.0	30.0	0.0	1636	0.689	1.073	0.0	1.0	SD
4	52	3.0	2.0	30.0	0.0	1321	0.827	1.329	1.0	0.0	SD

- Encoding de variáveis categóricas
- Normalização e padronização
- Engenharia de características polinomiais
- · Transformações logarítmicas
- Discretização (binning)
- · Feature scaling
- · Manipulação de outliers
- Tratamento de valores ausentes (imputação)
- · Seleção features
- Feature splitting
- Feature creating

Passo 2: Correlação



Estudo sobre a correlação (correlação de Pearson) entre as variáveis remanescentes e a variável alvo "CDR"

- 1.00

- 0.75

- 0.50

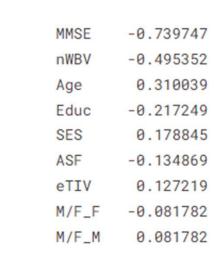
0.25

0.00

-0.25

-0.50

- -0.75



dataset[['MMSE', 'nWBV', 'Age', 'CDR']]



Descobrir insights escondidos nos dados

Diana Diaz, Data Scientist

Validação de hipóteses e modelos através de dados

Rafael Silva, Data Scientist

Passo 3: Hiperparâmetros

Métodos automatizados de busca de hiperparâmetros

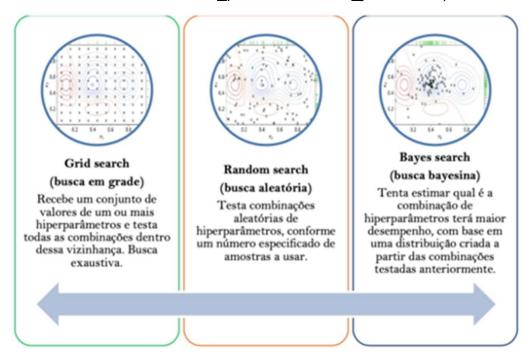


modelo = KNeighborsClassifier(n_neighbors = 3)
modelo.fit(X train, y train)

```
from sklearn.model_selection import GridSearchCV

param = {
        'n_neighbors': list(range(1, 41, 2)),
        'weights': ['uniform', 'distance'],
        'metric': ['euclidean', 'minkowski', 'manhattan'],
        'p': [1, 2]
}
knn = KNeighborsClassifier()
grid_search = GridSearchCV(knn, param, cv=3) # validação com 3 dobras
grid_search.fit(train.drop('CDR_', axis=1), train['CDR_'])
print("Melhores hiperparâmetros:")
print(grid_search.best_params_)
```

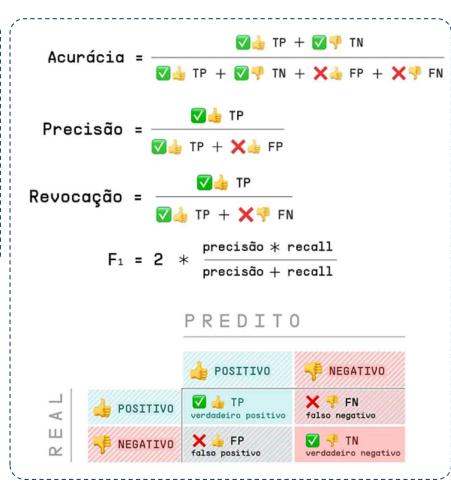
sklearn.neighbors.KNeighborsClassifier(n_neighbors=5, *, weights='uniform', algorithm='auto', leaf_size=30, p=2, metric='minkowski', metric_params=None, n_jobs=None)



Passo 4: Avaliação

Relatório de	Classificação precision		f1-score	support
SD	0.93	0.93	0.93	15
DML	0.75	0.60	0.67	5
DL	0.67	1.00	0.80	2
accuracy			0.86	22
macro avg	0.78	0.84	0.80	22
weighted avg	0.87	0.86	0.86	22

Verificação da qualidade do modelo



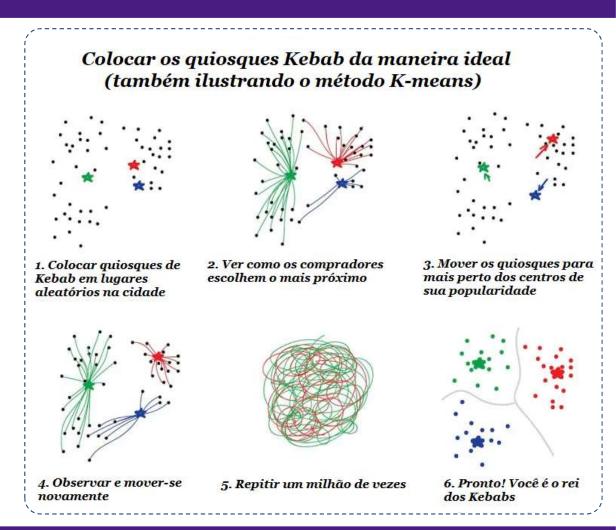
	Previsão	Esperado
28	SD	SD
206	SD	SD
189	SD	DML
37	SD	SD
169	SD	SD
38	DL	DL
151	DML	SD

Aprendizado não Supervisionado

 Não existem resultados pré-definidos para o modelo utilizar como referência para aprender

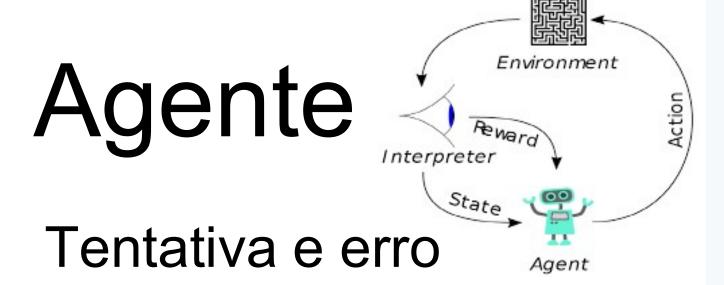
Exemplos de uso:

- Segmentação de mercado (tipos de clientes, fidelidade);
- Mesclar pontos próximos em um mapa;
- Compressão de imagem;
- Analisar e rotular novos dados;
- Detectar um comportamento anormal.



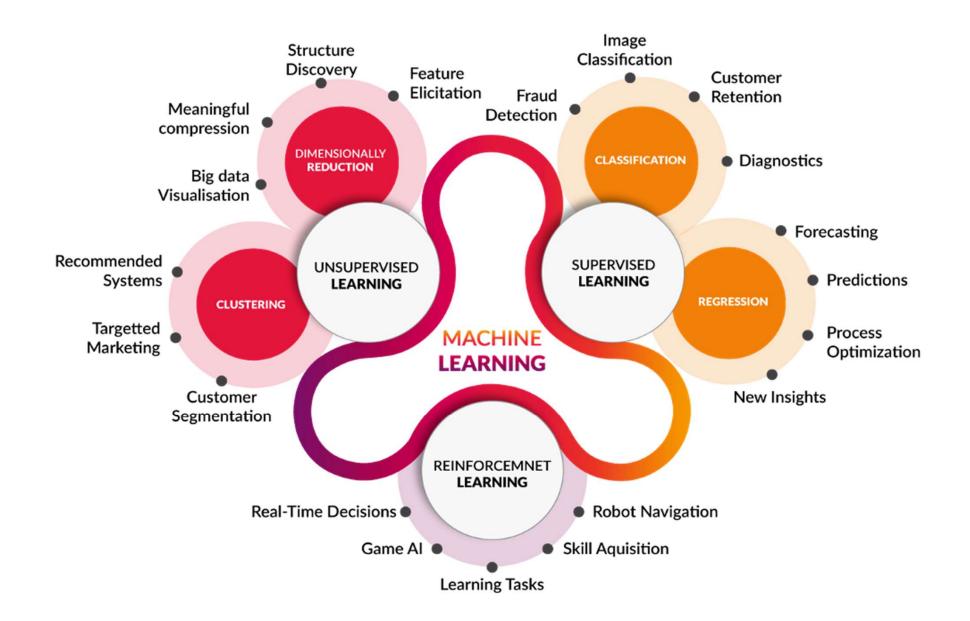


Aprendizado por Reforço



Recompensa ou punição

- Jogos manipuladores: jogos complexos como xadrez, Go e videogames. AlphaGo, desenvolvido pela DeepMind, é um exemplo famoso de agente RL que derrotou jogadores Go de classe mundial.
- Robótica: tarefas como braço robótico ao controle, navegação e direção autônoma.
- Sistemas de Recomendação:
 Empresas como Netflix use RL para melhorar seus algoritmos de recomendação. O sistema aprende com as interações e feedback dos usuários para sugerir conteúdo personalizado.
- Assistência médica: A RL é aplicada para otimizar planos de tratamento, dosagens de medicamentos e alocação de recursos médicos. Pode ajudar a personalizar as intervenções médicas para os pacientes.



Inteligência artificial (IA)

IA envolve técnicas que permitem que computadores emulem o comportamento humano, capacitando-os a aprender, tomar decisões, reconhecer padrões e resolver problemas complexos de maneira semethante à inteligência humana.

Machine learning (ML)

ML é um subconjunto da IA, utiliza algoritmos avançados para detectar padrões em grandes conjuntos de dados, permitindo que as máquinas aprendam e se adaptem. Os algoritmos de ML utilizam métodos de aprendizado supervisionado ou não supervisionado.

Deep learning (DL)

DL é um subconjunto do ML que utiliza redes neurais para o processamento de dados em profundidade e tarefas analibcas. O DL aproveita várias camadas de redes neurais artificiais para extrair características de alto nível a partir de dados brutos de entrada, simulando a maneira como os cérebros humanos percebem e compreendem o mundo.

IA generativa

A Inteligência Artificial Generativa é um subconjunto de modelos de Aprendizado Profundo que gera conteúdo, como texto, imagens ou código, com base em entradas fornecidas. Treinados em vastos conjuntos de dados, esses modelos detectam padrões e criam salidas sem instruções explicitas, utilizando uma combinação de aprendizado supervisionado e não supervisionado. Inteligência artificial Machine learning Deep learning IA generativa

Fonte: Zhuhadar, Lily Popova, Lytras, Miltiadis, The Application of AutoML Techniques in Diabetes Diagnosis: Current Approaches, Performance, and Future Directions, 2023/09/08.

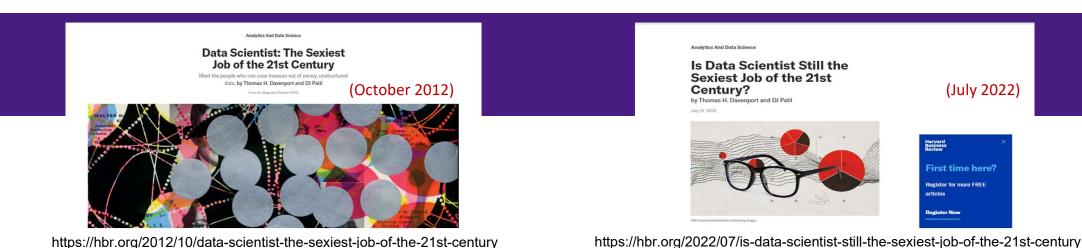
Artificial intelligence (AI):
Programs with the ability to
learn and reason like humans

Machine learning (ML):
Algorithms with the ability to learn
without being explicitly
programmed

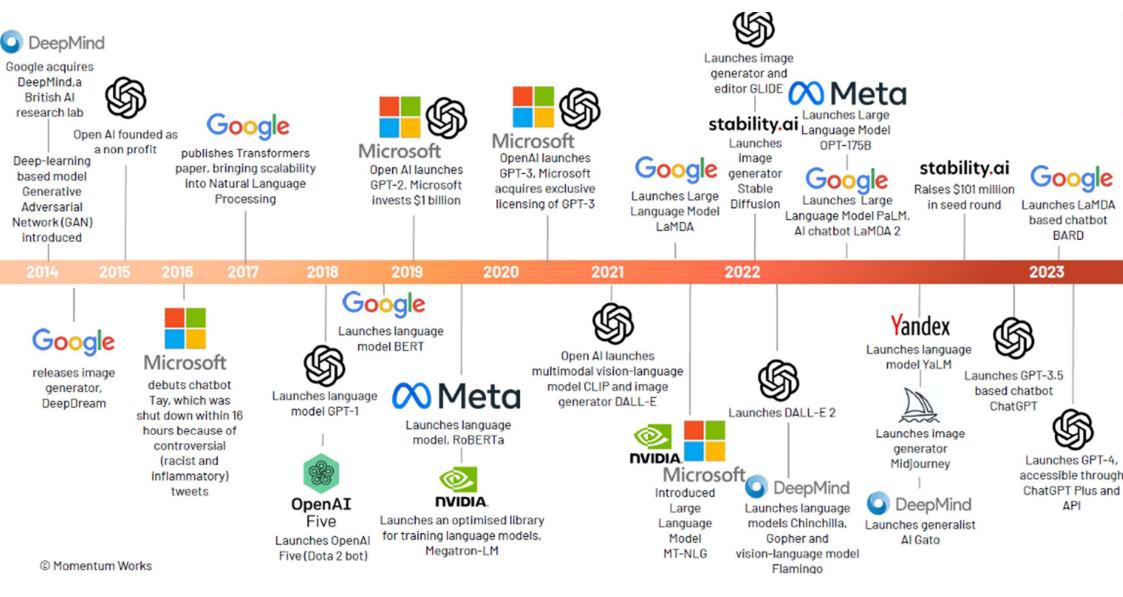
Deep learning (DL):
Subset of machine
learning in which artificial
neural networks adapt
and learn from large
datasets

Data Science:

A cross-disciplinary field that seeks to extract meaningful insights from data

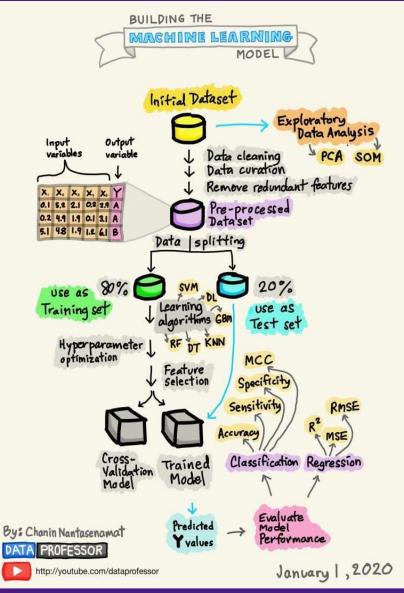


IBM Watson, um sistema de IA Bombe, máquina John McCarthy vence os John Hopfield cria Avanços no campo de criptografia cunha o termo competidores o conceito de redes de reconhecimento usada na Segunda Inteligência humanos neurais artificiais facial e tradução **Guerra Mundial** Artificial Alexa: assistente pessoal da 1950 1964 1997 2016 Amazon 1981 2014 1956 2017 Siri: assistente virtual de voz da ELIZA é **Apple** O supercomputador construído: AlphaGo derrota o Alan Turing propõe Deep Blue vence primeiro Chatbot campeão mundial o Teste de Turing Kasparov em uma do mundo com de Go partida de xadrez tecnologia NLP



Fonte: https://neigrando.com/2023/05/03/ia-generativa-um-relatorio-de-pesquisa-abril-2023/

Fonte: https://medium.com/@ai.insights/uma-breve-hist%C3%B3ria-da-evolu%C3%A7%C3%A3o-da-intelig%C3%AAncia-artificial-cb2b10f24e2c





https://www.linkedin.com/pulse/hiperpar%C3%A2metros-jose-r-f-junior/

Fonte das imagens: https://midia.market/conteudos/consumo/como-funciona-a-inteligencia-artificial/

A Beginner's Guide to The Machine Learning Workflow

datacaмр

Project setup

1. Understand the business goals

Speak with your stakeholders and deeply understand the business goal behind the model being proposed. A deep understanding of your business goals will help you scope the necessary technical solution, data sources to be collected, how to evaluate model performance, and more.

2. Choose the solution to your problem

Once you have a deep understanding of your problem—facus on which category of models drives the highest impact. See this Machine Learning Cheat Sheet for more information.

Data preparation

1. Data collection

Collect all the data you need for your models, whether from your own organization, public or paid sources.

2. Data cleaning

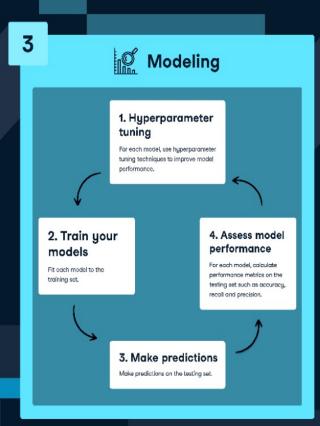
Turn the messy raw data into clean, tidy data ready for analysis. Check out this data cleaning checklist for a primer on data

3. Feature engineering

Manipulate the datasets to create variables (features) that improve your model's prediction accuracy. Create the same features in both the training set and the testing set.

4. Split the data

Randomly divide the records in the dataset into a training set and a testing set. For a more reliable assessment of model performance, generate multiple training and testing sets using cross-validation.



Deployment

1. Deploy the model

Embed the model you chose in dashboards, applications, or wherever you need it.

2. Monitor model performance

Regularly test the performance of your model as your data changes to avoid model drift.

3. Improve your model

Continously iterate and improve your model post-deployment. Replace your model with an updated version to improve performance.