































DADOS e APRENDIZAGEM AUTOMÁTICA

Conceitos e metodologias

MESTRADO (integrado) EM ENGENHARIA INFORMÁTICA

 #1 Artificial Intelligence AI / Machine Learning / Deep Learning	 #2 Internet of Things IOT, IIOT, Sensors & Wearables	 #3 Mobile/Social Internet Advancements - Search/Social/Messaging/Livestreams	 #4 Blockchain Distributed Ledger Systems, Cryptocurrencies & DApps	 #5 Big Data Apps, Infrastructure, Technologies + Predictive Analytics
 #6 Automation Information, Task, Process, Machine, Decision & Action	 #7 Robots Cons.,/Comm./Indus., Robots, Drones & Autonomous Vehicles	 #8 Immersive Media - #VR/ #AR/ #MR/ 360°/ Video?Gaming	 #9 Mobile Technologies Infrastructure, networks, standards, services & devices	 #10 Cloud Computing SaaS, IaaS, PaaS & MESH Apps
 #11 3D Printing Additive Manufacturing & Rapid Prototyping	 #12 CX Customer Journey, Experience Commerce & Personalization	 #13 EnergyTech Efficiency, Energy Storage & Decentralized Grid	 #14 Cybersecurity Security, Intelligence Detection, Remediation & Adaptation	 #15 Voice Assistants Interfaces, Chatbots & Natural Language Processing
 #16 Nanotechnology Computing, Medicine, Machines + Smart Dust	 #17 Collaborative Tech. Crowd, Sharing, Workplace & Open Source Platforms & Tools	 #18 Health Tech. Advanced Genomics, Bionics & Health Care Tech.	 #19 Human-Computer Interaction Facial/Gesture Recognition, Biometrics, Gaze Tracking	 #20 Geo-spatial Tech. GIS, GPS, Mapping & Remote Sensing, Scanning, Navigation
 #21 Advanced Materials Composites, Alloys, Polymers, Biomimicry, Nanomanufacturing	 #22 New Touch Interfaces Touch Screens, Haptics, 3D Touch, Paper, Feedback & Exoskeletons	 #23 Wireless Power Broadband incl. Fiber, 5G, Li-Fi, LPN and LoRa	 #24 Clean Tech. Bio-/Enviro-Materials + Solutions, Sustainability, Treatment & Efficiency	 #25 Quantum Computing + Exascale Computing
 #26 Smart Cities + Infrastructure & Transport	 #27 Edge/Computing + Fog Computing	 #28 Faster, Better Internet Broadband incl. Fiber, 5G, Li-Fi, LPN and LoRa	 #29 Proximity Tech Beacons, RFID, Wi-Fi, Near-Field Communications & Geofencing	 #30 New Screens TVs, Digital Signage, OOH, MicroLEDs & Projections

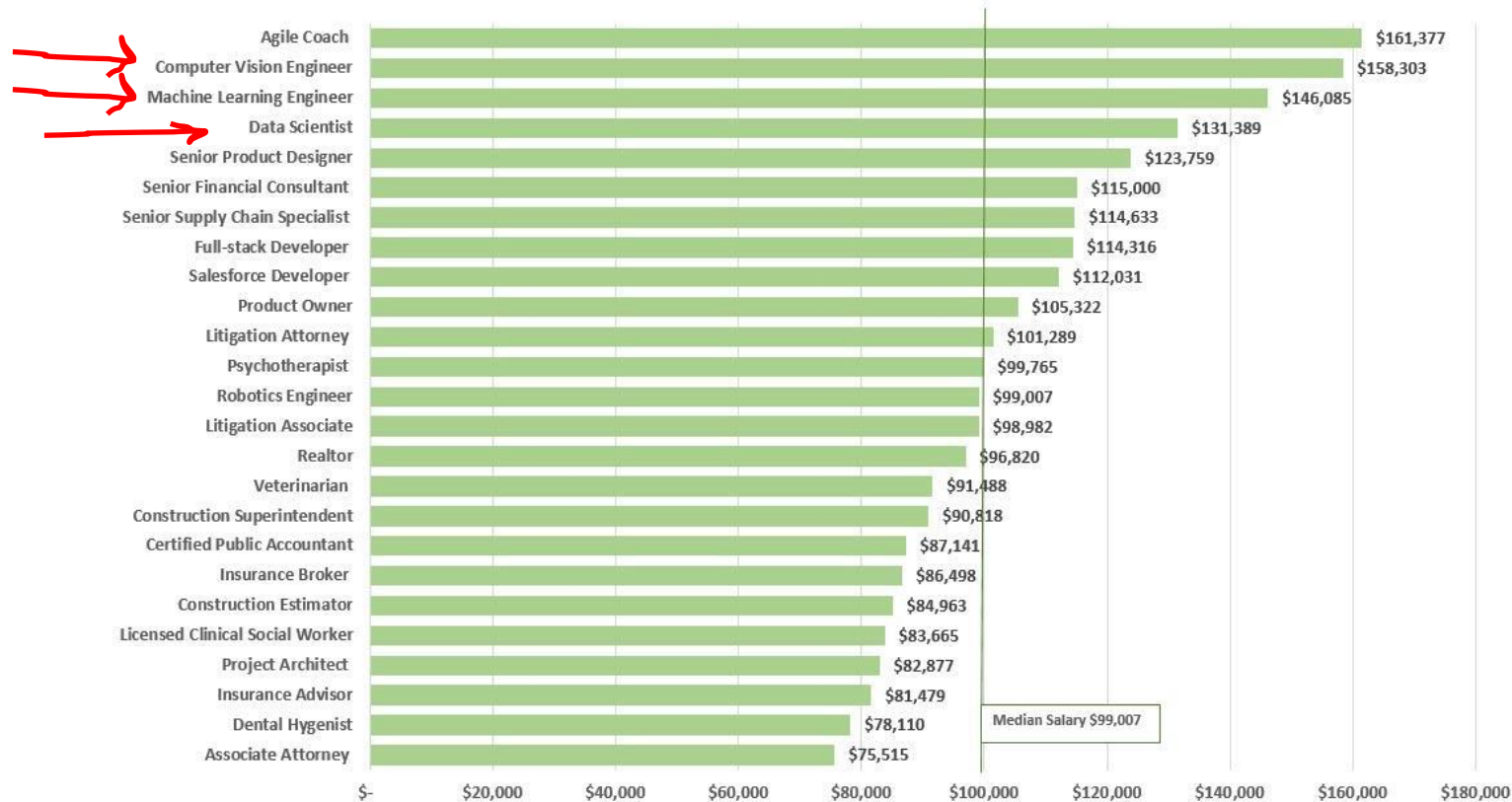
THE 30 TECHNOLOGIES OF THE NEXT DECADE

Created by: Sean Moffitt @seanmoffitt, Managing Director, @Wikibrands

CC BY NC SA

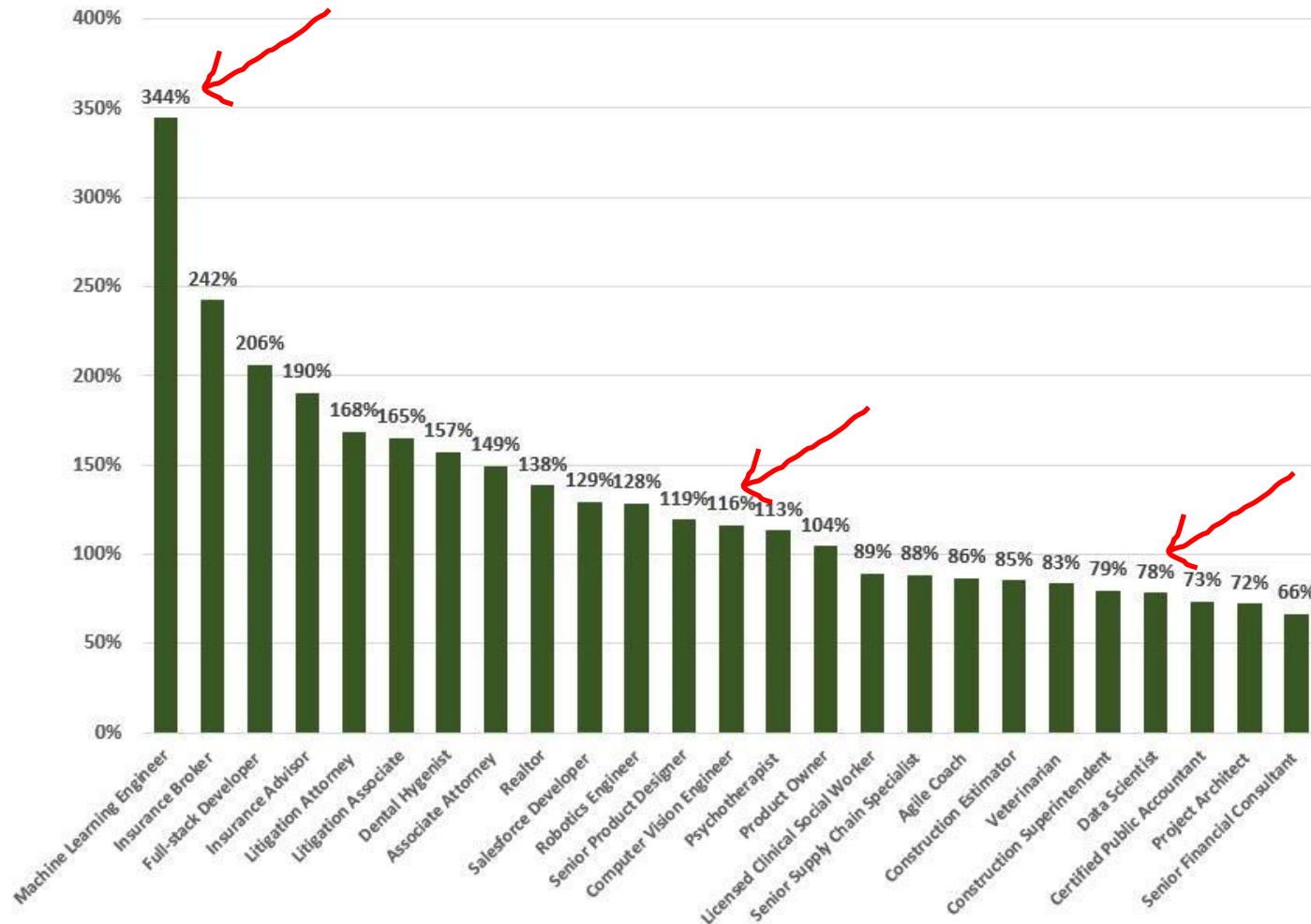
WIKIBRANDS

Indeed's Best Jobs In the U.S. Average Base Salary, 2019



Indeed's Best Jobs In The U.S.
% Growth in # of postings, 2015 - 2018
March 14, 2019

Motivação



Aprendizagem Automática vs Ciência de Dados

Aprendizagem Automática

vs

Ciência de Dados

(Machine Learning)

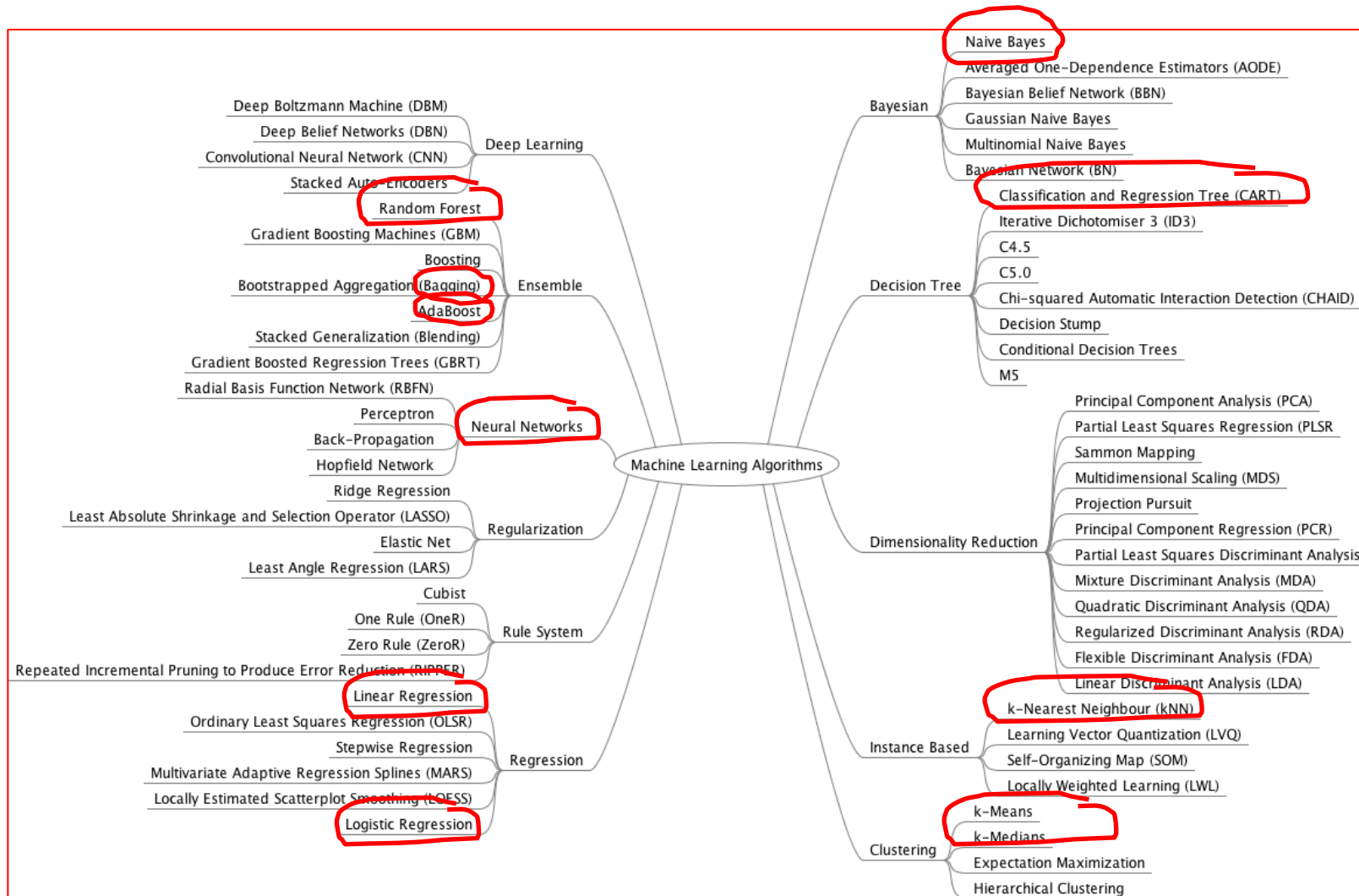
- Área que aborda as **ferramentas e técnicas de construção de modelos** que podem aprender por si próprios através da utilização de dados sem serem explicitamente programados.
- Utiliza dados
- O produto final é geralmente um **artefacto de software**
- Um engenheiro de aprendizagem automática (Data Engineer) **constrói modelos**.

(Data Science)

- Área científica que **estuda os dados** e como extrair **conhecimento e significado** dos mesmos
- Analisa conjuntos de dados
- O produto final é geralmente **apresentações e relatórios**;
- Um cientista de dados (data scientist) é um investigador que aplica as suas competências para elaborar uma **metodologia de investigação** e trabalha com a teoria por detrás dos algoritmos.

Não existe consenso universal sobre estes conceitos!!!

Algoritmos Aprendizagem Automática



“Ato ou efeito de aprender; Tempo durante o qual se aprende; Experiência que tem quem aprendeu.”

- Ganhar capacidades e conhecimentos
- Criar uma linha de raciocínio para resolver problemas
- Memorizar
- Reconhecer erros
- Corrigir erros
- Imitar comportamentos

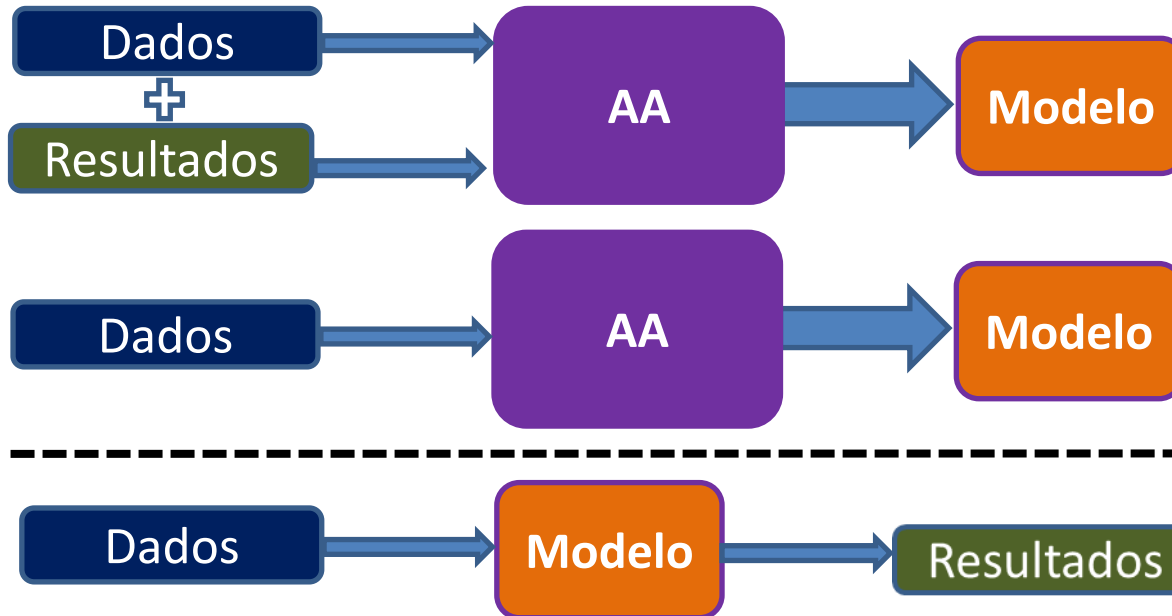
Paradigma de computação em que a característica essencial do sistema se revela pela sua **capacidade de aprender de modo autónomo e independente**.

- A característica diferenciadora dos algoritmos de Machine Learning é a de que são algoritmos **orientados aos dados**;
 - Um hipotético algoritmo aprenderia o que é um determinado objeto pela definição algorítmica da configuração desse objeto;
 - Um algoritmo de Machine Learning aprende sem necessidade de que seja codificada a solução do problema;
 - Um algoritmo de Machine Learning aprende a partir de diversos exemplos desse objeto, aprendendo desse modo a identificar esse objeto.

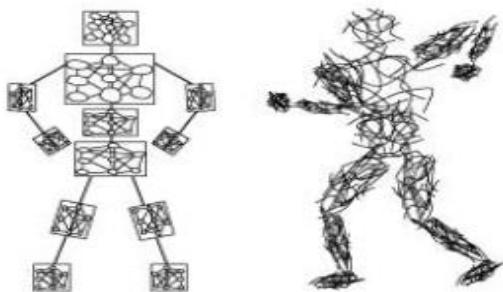
“Programação”



“Aprendizagem Automática”



Conceitos – Aprendizagem Simbólica vs não Simbólica



A **Aprendizagem Simbólica** (AS) refere-se ao facto de todos os passos se basearem em **representações simbólicas de leitura humana** do problema que utilizam a lógica e a procura para resolver o problema.

- A principal vantagem da AS é que o processo de raciocínio pode **ser facilmente explicado** - num programa de AS é fácil perceber porque é que se chega a uma determinada conclusão e quais foram os passos do raciocínio.
- Uma das desvantagens principais da AS é que, para o processo de aprendizagem – as **regras e o conhecimento precisam ser codificados à mão**, o que é um problema difícil.
- Até agora, a AS está muito confinada ao mundo académico e laboratórios universitários com pouco investimento dos gigantes da indústria.
- Uma das desvantagens principais da Aprendizagem não simbólica (AnS) é que é **difícil compreender** como é que o sistema chegou a uma conclusão. Isto é particularmente importante quando aplicado a aplicações críticas, tais como condução autónoma de automóveis, diagnóstico médico, entre outras.
- Em sistemas não simbólicos, como aplicações alimentados por DL, não são aceitáveis **decisões de alto risco**

Classifique os exemplos:

- Redes Neurais Artificiais;
- Raciocínio Baseado em Casos;
- Árvores de Decisão;
- Algoritmos Genéticos e Evolucionários;
- Máquinas de Vetores de Suporte;
- Inteligência de Grupo (swarm);
- Segmentação;
- Classificação;
- e muitos outros...

Paradigma de aprendizagem em que os casos que se usam para aprender incluem informação acerca dos resultados pretendidos, sendo possível estabelecer uma relação entre os valores pretendidos e os valores produzidos pelo sistema.

- A grande maioria dos algoritmos de Machine Learning usa aprendizagem supervisionada;
- Aprendizagem supervisionada significa que os dados de entrada (x) e os resultados (y), tornam possível que o algoritmo aprenda uma função (f) de mapeamento dos dados nos resultados: $y = f(x)$;
- Diz-se supervisionada porque este mapeamento é acompanhado por um algoritmo que supervisiona o processo de aprendizagem;
- Normalmente, são divididos em duas categorias:
 - **Classificação**: quando os resultados são discretos (e.g. preto, branco, cinza...);
 - **Regressão**: quando os resultados são contínuos (e.g. preço, temperatura idade,...).

Paradigma de aprendizagem em que não são conhecidos resultados sobre os casos, apenas os enunciados dos problemas, tornando necessário a escolha de técnicas de aprendizagem que avaliem o funcionamento interno do sistema.

- A aprendizagem não supervisionada significa que existem dados de entrada (x) mas não existem os correspondentes resultados;
- O objetivo deste tipo de aprendizagem é o de modelar a estrutura ou a distribuição dos dados do problema;
- São, normalmente, divididos em duas categorias:
 - **Segmentação (clustering)**: quando se pretende organizar os dados em grupos coerentes (agrupar clientes que comprem produtos biológicos);
 - **Redução (reduction)**: reduzir o número de características de um conjunto de dados ou decompor o conjunto de dados em múltiplos componentes;
 - **Associação**: quando se pretende conhecer regras que associem o comportamento demonstrado pelos dados (pessoas que comprem produtos biológicos não comprem produtos de charcutaria):

Paradigma de aprendizagem que, apesar de não ter informação sobre os resultados pretendidos, permite efetuar uma avaliação sobre se os resultados produzidos são bons ou maus.

- Algoritmos de Reinforcement Learning usam técnicas de auto-alimentação de sinais, com vista a melhorar os resultados, por influência da noção de recompensa/penalização;
- Não se pode comparar com Aprendizagem Supervisionada uma vez que a “avaliação” dos resultados não é dada por um supervisor;
- Também não se pode considerar Aprendizagem não Supervisionada, uma vez que não existe ausência absoluta de informação sobre os resultados;
- A aprendizagem dá-se pela capacidade de crítica sobre os próprios resultados produzidos pelo algoritmo;
 - **Q-Learning**: assume que está a seguir uma política ótima e usa-a para atualização dos valores das ações;
 - **SARSA**: considera a política de controlo que está a ser seguida e atualiza o valor das ações.

Classifique cada um destes casos quanto ao tipo de aprendizagem:

- | | |
|--------------------------------|------------------------------|
| 1. Classificação de imagem | 11. Fidelização de clientes |
| 2. Diagnóstico | 12. Detecção de fraude |
| 3. Aquisição de aptidões | 13. Navegação de robôs |
| 4. Decisões em tempo real | 14. Tarefas de aprendizagem |
| 5. Jogos com IA | 15. Crescimento populacional |
| 6. Previsão de mercados | 16. Previsão meteorológica |
| 7. Esperança média de vida | 17. Visualização (Big Data) |
| 8. Compreensão de significados | 18. Descoberta de Estruturas |
| 9. Seleção de atributos | 19. Segmentação de clientes |
| 10. Marketing | 20. Sistemas de recomendação |



<https://forms.gle/SRqdN98P1BAotif57>

Aprendizagem Automática (Machine Learning)

Aprendizagem supervisionada (Supervised Learning)

- Classificação (Classification)

- Classificação de imagem
- Fidelização de clientes
- Diagnóstico
- Detecção de fraude

- Regressão (Regression)

- Crescimento populacional
- Previsão de mercados
- Previsão meteorológica
- Esperança média de vida

Aprendizagem não supervisionada (Unsupervised Learning)

- Redução (Reduction)

- Visualização (Big Data)
- Compreensão de significados
- Descoberta de Estruturas
- Seleção de atributos

- Segmentação (Clustering)

- Segmentação de clientes
- Marketing
- Sistemas de recomendação

Aprendizagem por reforço (Reinforcement Learning)

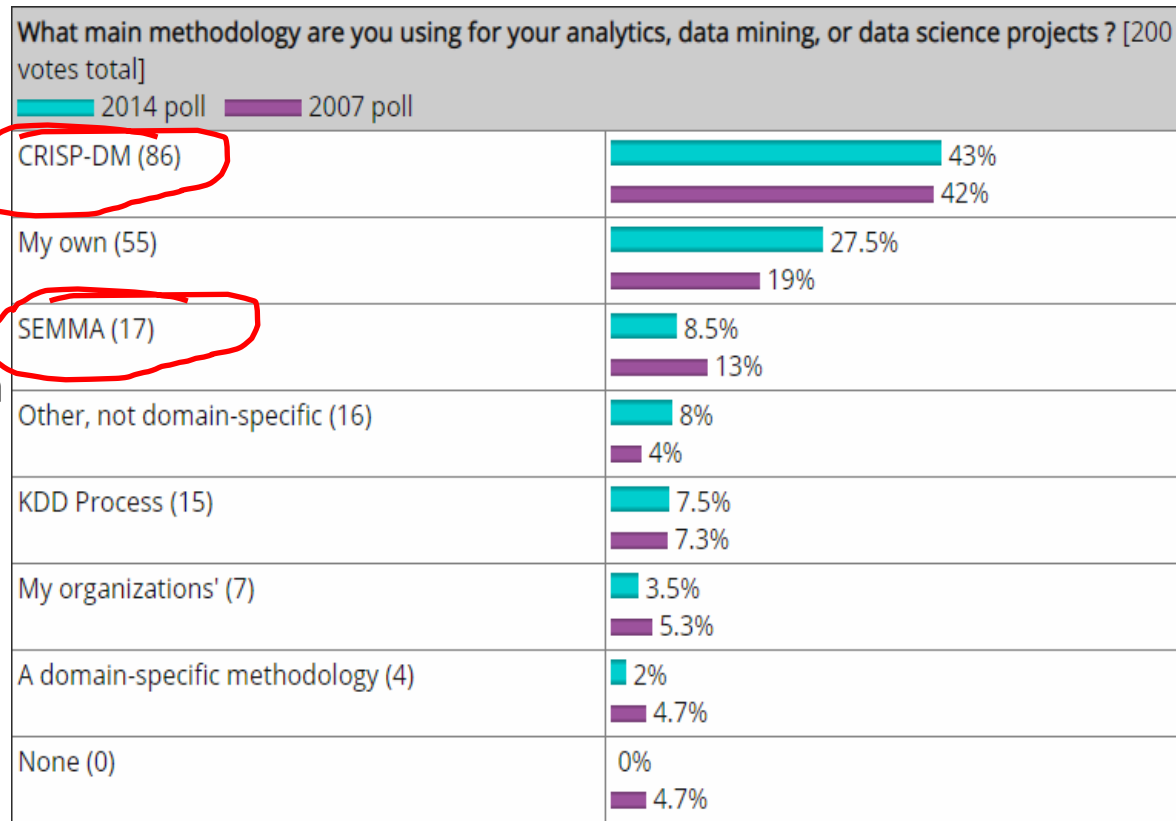
- Aquisição de aptidões
- Navegação de robôs
- Decisões em tempo real
- Tarefas de aprendizagem
- Jogos com IA

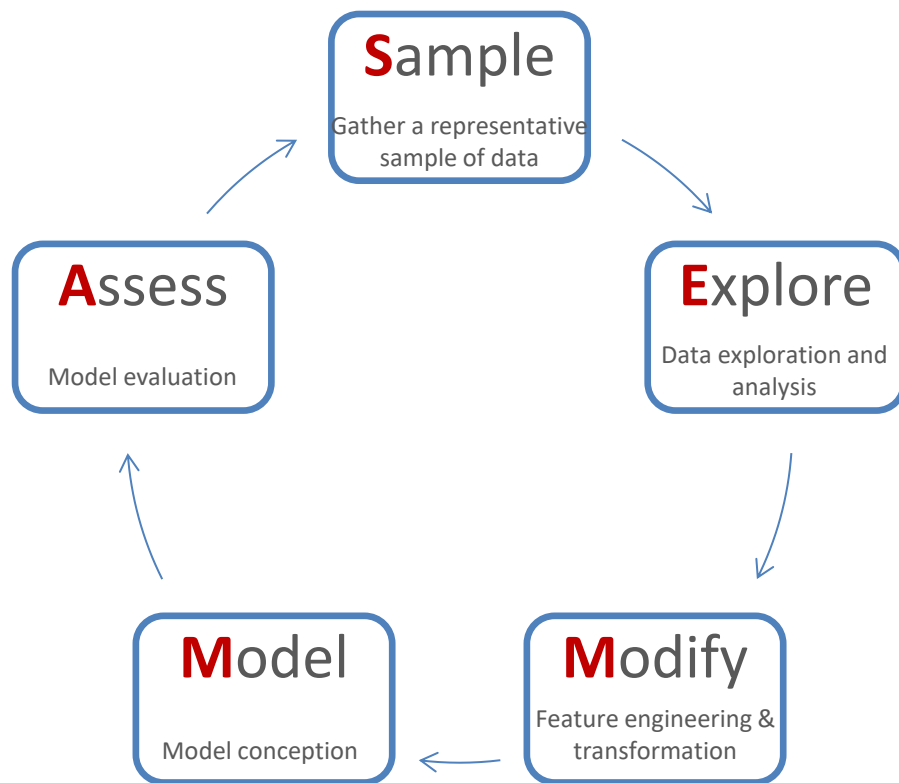
Metodologias

Data Science Pipeline

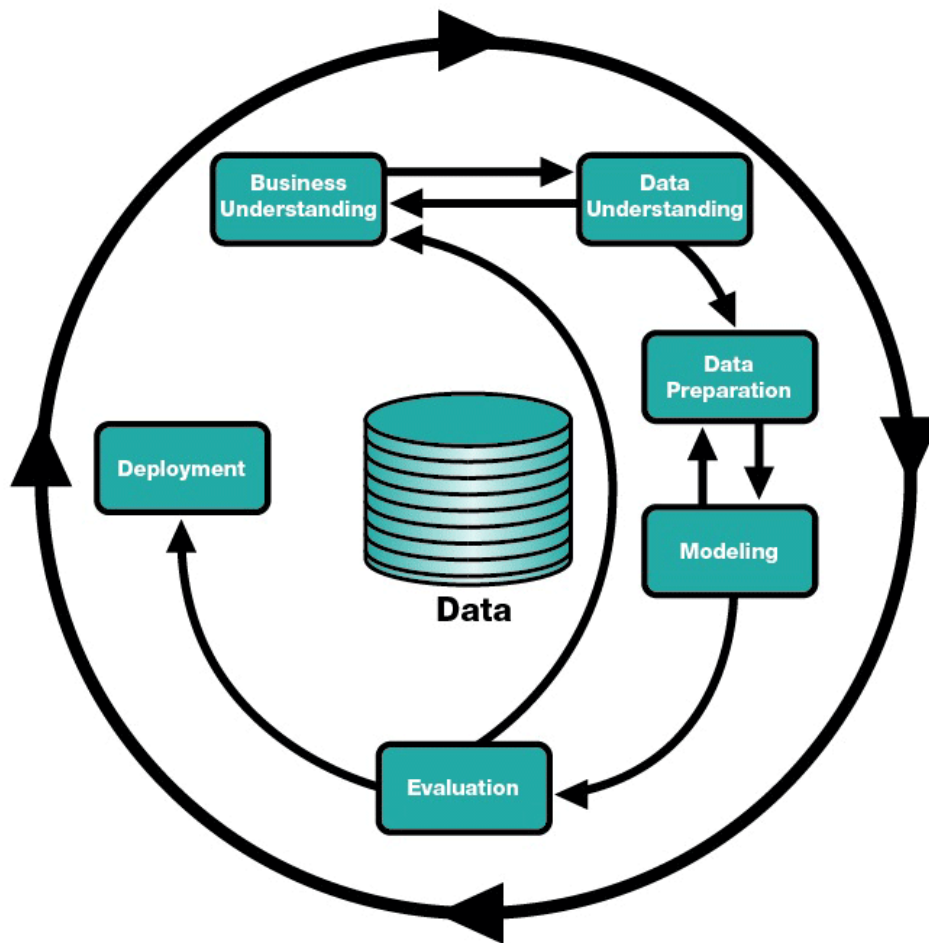
Motivos para utilizar uma metodologia:

- Permite que os projetos sejam **replicados**;
- Apoia no **planeamento e gestão do projeto**;
- Incentiva as melhores práticas e ajuda a obter **melhores resultados**.



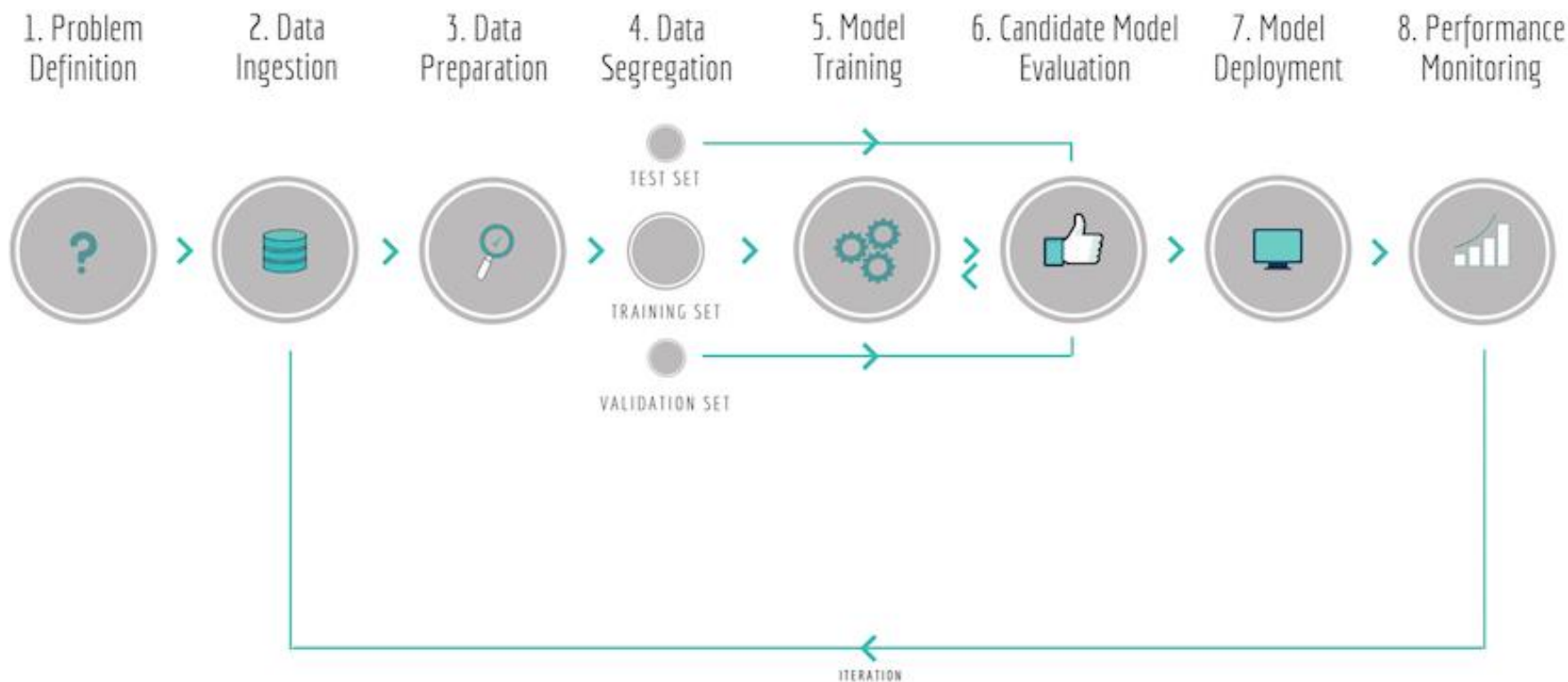


Cross Industry Standard Process for Data Mining (**CRISP-DM**)



Fluxo Aprendizagem Automática

Fluxo de um processo de Aprendizagem Automática



DADOS e APRENDIZAGEM AUTOMÁTICA

Conceitos e metodologias

MESTRADO (integrado) EM ENGENHARIA INFORMÁTICA