

IA e ML Aplicados a Finanças

Prof. Leandro Maciel

AULA 1: Sistemas Inteligentes e Aprendizagem

- 1 Motivação
- 2 Sistemas Inteligentes e IA
- 3 Machine Learning
- 4 Aprendizagem
- 5 Bibliografia

1. Motivação

- Mundo atual: **“inundação de dados”**;
- International Data Corporation (IDC):
 - 163 zettabytes (10^{21}) de dados gerados no mundo até 2025;
 - contra 4,4 zettabytes produzidos até 2013.
- **Internet of Things (IoT)**:
 - interconexão digital de objetos cotidianos com a Internet;
 - conexão maior do que das pessoas (metaverso);
 - uber, google, whatsapp, instagram, exposições “imersivas”.

Internet (FORBES, 2018):

- Mais da metade pesquisas são realizadas em *smartphones*;
- 3,7 bi de pessoas usam internet (aumento 7,5% em relação 2016);
- Google processa mais de 40.000 buscas **cada segundo**;
- 77% pesquisas são realizadas no Google;
- 5 bilhões de pesquisas são realizadas diariamente.

Mídias Sociais (FORBES, 2018):

- Grande combustível para criação de dados. A cada minuto:
- Usuários do Snapchat compartilham 527.670 fotos;
- Mais de 120 profissionais aderem ao LinkedIn;
- 4.146.600 vídeos são visualizados no YouTube;
- 456.000 mensagens são compartilhadas no Twitter;
- 46.740 fotos são postadas no Instagram.

Facebook (FORBES, 2018):

- Diariamente 1,5 bi de pessoas estão ativas na rede social;
- Mais de 307 mi de pessoas são na Europa;
- 5 perfis no Facebook são criados a cada segundo;
- Mais de 300 mi de fotos são carregadas diariamente;
- Cada minuto 510.000 comentários e 293.000 atualizações de status.

Comunicação a cada minuto:

- 16 mi de mensagens de texto;
- 990.000 “deslizar” no Tinder;
- 156 mi e-mails são enviados;
- 15.000 GIFs são enviados via Facebook;
- 103.447.530 *spams* são enviados via e-mail.

■ Cenário atual:

- *smartphones* realizam interações e coletam dados;
- explosão de dados (**Big Data**);
- como dar sentido e lógica para essas informações?
- dados como instrumento de **vatagem competitiva**;
- cada vez mais dados irão requerer análise;
- inovação baseada em dados → **Data-Driven**.

■ Ferramentas: **Data Science** e **Data Analytics**.

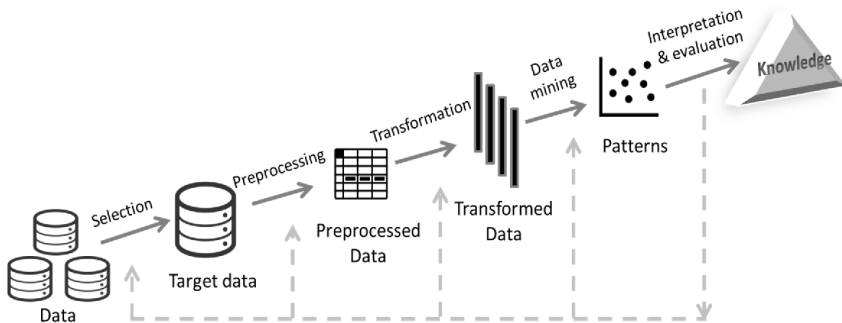
■ Data Science ou Ciência de Dados:

- des. tecnologias e métodos para obter, organizar e processar infos.;
- **extração de conhecimento** dos dados;
- objetivo de agregação de valor ao negócio;
- preocupação com armazenamento e segurança.

■ Data Analytics ou Análise de Dados:

- faz parte da ciência de dados;
- soluções, padrões, tendências para tomada de decisão.

- Processos na ciência de dados:

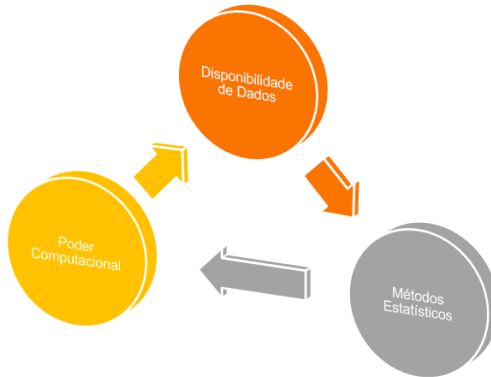


- Como realizar a análise de dados?
- Métodos:
 - modelos matemáticos e estatísticos, de ML e IA.
- Ferramentas:
 - *softwares* estatísticos e de programação.
- Requisitos: matemática, estatística, otimização e programação.

- Ferramentas para análise e processamento de dados:



- Motivadores do desenvolvimento de técnicas de IA e ML:



2. Sistemas Inteligentes e IA

- Sistema inteligente tem as seguintes capacidades:
 - raciocinar, planejar, resolver problemas, armazenar conhecimento, adaptar-se, aprender, etc.

- Definição de **Inteligência Artificial (IA)**:
 - campo da ciência da computação e da engenharia de computação que procura reproduzir, por meios computacionais, as capacidades de sistemas inteligentes;
 - habilidade de obter sucesso em novas situações pelo ajuste do comportamento (**processamento inteligente da informação**);
 - *algoritmos computacionais capazes de pensar, sentir e agir.*

■ Nascimento da IA:

- termo IA cunhado em um workshop no Dartmouth College (Hanover, New Hampshire), em 1956, por John McCarthy;
- humanidade sempre interessada por responder questões envolvendo a inteligência nos mais variados contextos.

■ Histórico da IA...

■ Anos 40:

- McCulloch e Pitts (1943) criam primeiro modelo matemático de um neurônio;
- ENIAC (Electronic Numerical Integrator and Computer) → criado em 1946, sendo o primeiro computador digital eletrônico de grande escala
- Norbert Wiener cunhou, em 1948, o termo cibernética, voltado para todo dispositivo que realizasse conjuntamente controle e processamento de informação;
- Regra de Hebb (1949) para ajuste de pesos de redes neurais artificiais.

■ Anos 50:

- Dartmouth Summer Research Project on Artificial Intelligence (1956);
- Estratégias de busca (jogos de tabuleiro, busca de alvo em labirintos);
- Multilayer perceptron (Rosenblatt, 1958) → primeiro modelo (classificador binário linear) para aprendizado supervisionado.

■ Anos 60:

- GPS (General Problem Solver) (Newell & Simon, 1961);
- Robótica, sistemas baseados em regras;
- Lógica nebulosa e algoritmos evolutivos.

■ Anos 70:

- representação do conhecimento;
- sistemas especialistas;
- complexidade computacional.

■ Anos 80:

- aprendizagem e adaptação (RNAs);
- aplicações industriais.

■ Anos 90:

- inteligência computacional;
- aprendizado de máquina;
- mineração de dados;
- agentes e sistemas multiagentes.

■ Século 21:

- tecnologia do relacionamento (lei de Moore e comunicação sem fio);
- ChatGPT como protótipo de um chatbot com inteligência artificial desenvolvido pela OpenAI especializado em diálogo.

3. Machine Learning

- Objetivos das mais diversas áreas (finanças):
 - aprendizagem de mapeamentos, relações, causalidade...
 - **dados** como fonte de informação;
 - via modelos com características diversas (+ ou – restritivos).

- O que é machine learning (ML)?
 - algoritmos/métodos e modelos estatísticos computacionais que “aprendem” padrões escondidos em dados;
 - predição, classificação, reconhecimento padrões, agrupamento, etc.

■ ML:

- inferência e análise causal não é fundamental;
- modelos interpretáveis não é crucial;
- ML com propriedades estatísticas → **Statistical Learning**.

■ Econometria/Econometria Financeira:

- métodos estatísticos para previsão inferência, modelagem causal de relações econômicas/financeiras;
- inferência é crucial e interpretação muito relevante;
- causalidade desempenha papel central na tomada de decisões.

■ Big Data:

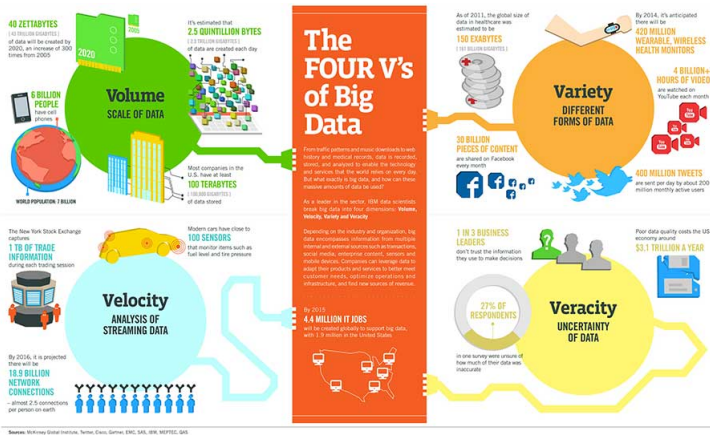
- grande quantidade de dados (variáveis/observações);
- negócios captam dados diariamente;
- futuro das organizações depende do uso adequado dos dados;
- melhores decisões estratégicas.

■ Dados estruturados e não estruturados, respectivamente:

- organizados (estrutura de base de dados, fácil acesso);
- todo o resto (textos, páginas web, mídia social, e-mail, etc).

3. Machine Learning

- Os 4 “V’s” do Big Data (<http://www.ibmbigdatahub.com>):



- Componentes no aprendizado de máquinas:
 - **conjunto de dados**, a ser particionado;
 - **modelo matemático** (síntese de um mapeamento multidimensional de entrada-saída), flexível e com parâmetros livres;
 - **função perda ou custo**, que mensura o desempenho do modelo;
 - **procedimento de otimização** (aprendizagem), com objetivo de minimizar função custo.

- Em um modelo de regressão linear simples, quais são esses componentes?

- ML em finanças empíricas:
 - previsão prêmio pelo risco;
 - previsão de retornos;
 - determinantes dos retornos;
 - estimação covariâncias e seleção de carteiras;
 - modelagem e previsão da volatilidade;
 - previsões macroeconômicas;
 - score de crédito, detecção de fraudes, trading algorítmico.

3. Aprendizagem

- Aprendizagem em ML → ajuste de parâmetros livres;
- Aprendizado supervisionado:
 - uso de dados **rotulados**;
 - saída desejada apresentada ao modelo.
- Aprendizado não-supervisionado:
 - uso de dados **não-rotulados**;
 - clusterização (similaridade) e função densidade de probabilidades.
- Aprendizado por reforço:
 - recompensa ou punição por decisões sequenciais.

- Seremos substituídos por máquinas que aprendem com dados?
- Etapas do processo de aprendizado:
 - entender;
 - usar/praticar conceitos;
 - explicar para outras pessoas;
 - transpor para outros contextos;
 - criticar/detectar limitações e extensões possíveis (propor inovações).
- As máquinas possuem essas capacidades?

- *De aorcdõ com uma pqsieusa de uma uinrvsriddae ignlsea , não ipomtra em qaul odrem as lrtas de uma plravaa etãso , a úncia csioa iprotmatne é que a piremria e útmliã lrtas etejasm no lgaur crteo . O rseto pdoe ser uma ttaol bçguana que vcoê pdoe andia ler sem pborlmea . Itso oorcre pporue nós não lmeos cdaa lrtã isladoa, mas a plravaa cmoo um tdoo*
- Você consegue ler esse texto?
- Uma máquina consegue ler esse mesmo texto?

- Google Translate (6 de fev. de 2023):
 - *According to a pqsieusa of a uinrvsriddae ignlsea, not ipomtra in qaul odrem the lines of a plravaa etão , the uncia csioa iprotmatne is that pyremria and útmlia l rteas etejasm no lgaur crteo . O rseto can be a ttaol bçguana that you are pdoe andia ler sem pborlmea . It's gold why we don't film cdaa l rtea isladoa , but the plravaa cmoo um todoo*
- Por que a máquina não se deu tão bem nessa tarefa?

4. Aprendizagem

- Encontre o “5” que fica acima do “+”, à direita do “vazio”, abaixo do “3” e à esquerda do “@”:

	1	2	3	4	5	6	7	8	9	10	11	12
a												
b												
c												
d												
e												
f												
g												
h												
i												
j												
k												
l												

- Encontre o “5” que fica acima do “+”, à direita do “vazio”, abaixo do “3” e à esquerda do “@”:

	1	2	3	4	5	6	7	8	9	10	11	12
a	@			5	@	+		3		5		+
b	+	5	+		3		5	@	5	3	@	
c	3			@	5	@	+	5			5	
d		+		5		5	+		+		3	+
e	+	5	3	+	3	+		5		3		@
f		+	5		5	@		+		5		3
g	3		3	5	+		5	3	@	+		5
h	5		@		5		3		3		3	
i	3		5	3	+	5	@	+	5			+
j		+	+		@		5				3	5
k	@	3		5	+	5	3	@		5	+	
l	5		@	+			+		3	+		5

4. Aprendizagem

■	3	■
■	5	@
■	+	■

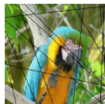
	1	2	3	4	5	6	7	8	9	10	11	12
a												
b												
c												
d												
e												
f												
g												
h												
i												
j												
k												
l												

4. Aprendizagem

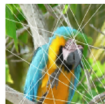
■	3	■
■	5	@
■	+	■

	1	2	3	4	5	6	7	8	9	10	11	12
a	@			5	@	+		3		5		+
b	+	5	+		3		5	@	5	3	@	
c	3			@	5	@	+	5			5	
d		+		5		5	+		+		3	+
e	+	5	3	+	3	+		5		3		@
f		+	5		5	@		+		5		3
g	3		3	5	+		5	3	@	+		5
h	5		@		5		3		3		3	
i	3		5	3	+	5	@	+	5			+
j		+	+		@		5				3	5
k	@	3		5	+	5	3	@		5	+	
l	5		@	+			+		3	+		5

■ Degradação de imagem e identificação:



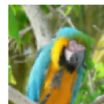
(a) Black lines



(b) White Lines



(g) Global Blur



(h) Local Blur



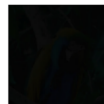
(c) Small Boxes



(d) Random Noise



(i) White Fog



(j) Fade to Black



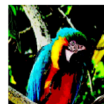
(e) Adj. Pixel Swap



(f) Rand. Pixel Swap



(k) Fade to white



(l) Posterise

■ Formulação:

- problema precisa ser bem formulado;
- facilita a tarefa de aprendizagem.

■ Adequação:

- sistemas (humano/máquina) têm diferentes habilidades;
- explorar habilidades mais proeminentes;
- classificadores, preditores, regressores.

- Não podemos esquecer de aspectos éticos em IA:



Stanford Encyclopedia of Philosophy

Ethics of Artificial Intelligence and Robotics

First published Thu Apr 30, 2020

Artificial intelligence (AI) and robotics are digital technologies that will have significant impact on the development of humanity in the near future. They have raised fundamental questions about what we should do with these systems, what the systems themselves should do, what risks they involve, and how we can control these.

After the Introduction to the field (§1), the main themes (§2) of this article are: Ethical issues that arise with AI systems as *objects*, i.e., tools made and used by humans. This includes issues of privacy (§2.1) and manipulation (§2.2), opacity (§2.3) and bias (§2.4), human-robot interaction (§2.5), employment (§2.6), and the effects of autonomy (§2.7). Then AI systems as *subjects*, i.e., ethics for the AI systems themselves in machine ethics (§2.8) and artificial moral agency (§2.9). Finally, the problem of a possible future AI superintelligence leading to a “singularity” (§2.10). We close with a remark on the vision of AI (§3).

- Fonte: <https://plato.stanford.edu/entries/ethics-ai/>.

■ Próxima aula...

- reforço do ferramental básico;
- programação e modelo ML mais simples (regressão linear)...

Dixon, M. F., HALPERIN, I. & BILOKON, P. **Machine Learning in Finance: From Theory to Practice**. Springer, 2021.

Fetzer, J. H. **Computers and Cognition: Why Minds are not Machines**. Dordrecht: Kluwer Academic Publishers, 2001.

Russell, S. & Norvig, P. **Artificial Intelligence - A Modern Approach**. 4th Edition, Pearson, 2020.

TURING, A.M. Computing machinery and intelligence, In Mind - A Quarterly Review of Psychology and Philosophy, Vol. LIX No. 236, pp. 433-460, Oct. 1950.

Prof. Leandro Maciel

leandromaciel@usp.br