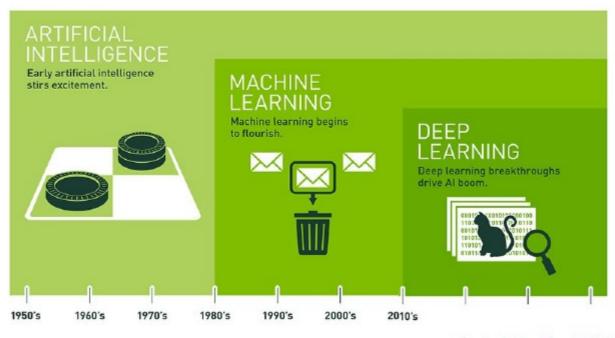
Aprendizado de Máquina ou Machine Learning

Prof. Marcelo Dib Cruz

Artigo

Aprendizado de Máquina

Perspectiva histórica



[fonte: Artigo Blog NVidia]



Aprendizagem

- Aprendizado supervisionado
 - Regressão
 - Classificação
- Aprendizado não supervisionado
 - Clusterização
 - Redução de dimensão
 - Detecção de anomalias
- Aprendizado por reforço
 - Processo de decisão usando Markov



- Aprendizado Supervisionado
 - Classificação
 - Árvore de Decisão
 - Regressão
 - Regressão Linear
 - Redes Neurais
- Aprendizado n\u00e3o supervisionado
 - RNA não supervisionada
 - Clustering ou Agrupamento
 - K-means
 - Automatic Clustering
 - Algoritmo Genético
- Aprendizado por Reforço

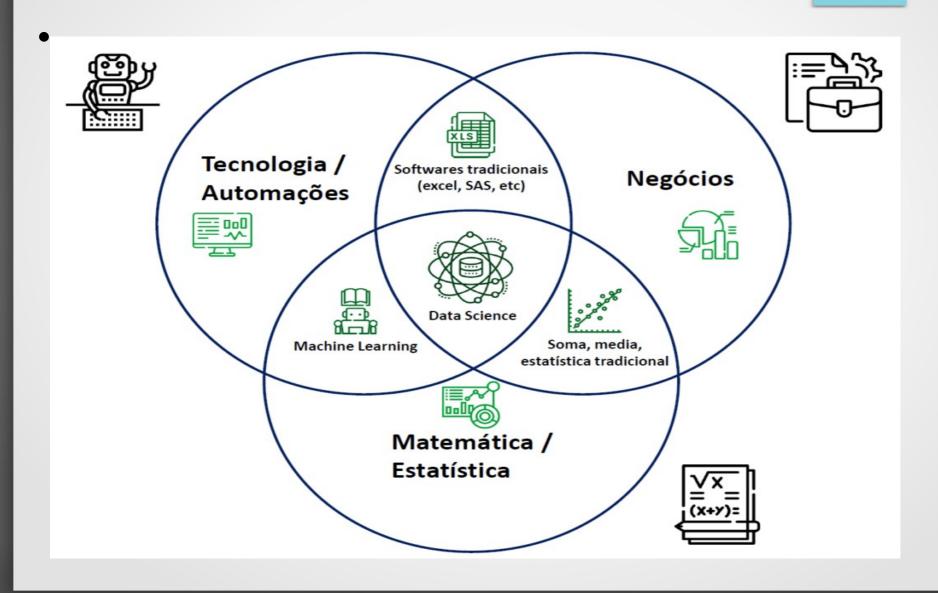
- Workflow típico para resolução de problemas
 - Definição do problema
 - Carregamento do dataset
 - Análise e visualização dos dados
 - Preparação dos dados
 - Limpeza dos dados
 - Transformação dos dados
 - Geração e treinamento dos modelos
 - Avaliação dos modelos
 - Apresentar os resultados (tabelas, gráficos)

Ciência de Dados

- A ciência de dados combina vários campos, incluindo estatísticas, métodos científicos, inteligência artificial (IA) e análise de dados para extrair valor dos dados. Aqueles que praticam ciência de dados são chamados de cientistas de dados e combinam uma variedade de habilidades para analisar dados coletados da Web, de smartphones, clientes, sensores e outras fontes para obter insights acionáveis.(Site da Oracle)
- A ciência de dados abrange a preparação de dados para análise, incluindo limpeza, agregação e manipulação dos dados para realizar análises avançadas de dados. As aplicações analíticas e os cientistas de dados podem revisar os resultados para descobrir padrões e permitir que os líderes de negócios obtenham insights informados.
- Data Science is different from other areas such as mathematics or statistics.
 Data Science is an applied activity and data scientists serve the needs and solve the problems of data users. Before you can solve a problem, you need to identify it and this process is not always as obvious as it might seem.

Ciência de Dados

- Ela é usada nas organizações para ajudar a melhorar seu funcionamento e a criar valor.
- A ciência de dados possui ferramentas para analisar esses dados e a partir deles, extrair conhecimentos essenciais para processos de tomada de decisão nos mais diversos contextos, como:
 - estratégias de logística
 - melhoria do trânsito urbano
 - controle de estoque de produtos
 - alocação de tarefas a colaboradores
 - melhoria de desempenho de alunos, atletas e vendedores
 - redução de custos
 - detecção de fraudes



Exemplo:

Problema de previsão de Empréstimo



*DESCRIÇÃO DAS VARIÁVEIS:

Variável Descrição

Loan_ID
 ID único do empréstimo (loan)

Gender Masculino/Feminino (Male/Female)

Married Casado - sim ou não (Y/N)

Dependents
 Numero de dependentes

Education
 Grau escolar (Graduate/ Under Graduate)

Self_Employed Auto empregado - sim ou n\u00e4o (Y/N)

ApplicantIncome Renda do aplicante

- CoapplicantIncome Renda do co-aplicante

LoanAmount
 Montante do empréstimo (loan), em milhares

- Loan Amount Term Prazo do empréstimo (loan), em meses

Credit_History
 Histórico de crédito corresponde aos critérios

Property_Area Localização da propriedade (Urban/Semi Urban/ Rural)

Loan_Status
 Empréstimo (loan) aprovado - sim ou não (Y/N)

Loan ID	Gender	Married	Dependen	Education	Self Employe	ApplicantInco	CoapplicantIn
LP001002	Male	No		Graduate	No	5849	0
LP001003	Male	Yes		Graduate	No	4583	1508
LP001005	Male	Yes		Graduate	Yes	3000	0
LP001006	Male	Yes		Not	Graduate	No	2583
LP001008	Male	No		Graduate	No	6000	0
LP001011	Male	Yes	2	Graduate	Yes	5417	4196
LP001013	Male	Yes	0	Not	Graduate	No	2333
LP001014	Male	Yes	3+	Graduate	No	3036	2504
LP001018	Male	Yes	2	Graduate	No	4006	1526
LP001020	Male	Yes	1	Graduate	No	12841	10968
LP001024	Male	Yes	2	Graduate	No	3200	700
LP001027	Male	Yes	2	Graduate		2500	1840
LP001028	Male	Yes	2	Graduate	No	3073	8106
LP001029	Male	No	0	Graduate	No	1853	2840
LP001030	Male	Yes	2	Graduate	No	1299	1086
LP001032	Male	No	0	Graduate	No	4950	0
LP001034	Male	No	1	Not	Graduate	No	3596

LoanAmount	Loan_Amount	Credit_History	Property_Area	Loan_Status	
	360	1	Urban	Υ	
128	360	1	Rural	Ν	
66	360	1	Urban	Υ	
2358	120	360	1	Urban	Υ
141	360	1	Urban	Υ	
267	360	1	Urban	Υ	
1516	95	360	1	Urban	Υ
158	360	0	Semiurban	Ν	
168	360	1	Urban	Υ	
349	360	1	Semiurban	Ν	
70	360	1	Urban	Υ	
109	360	1	Urban	Υ	
200	360	1	Urban	Υ	
114	360	1	Rural	Ν	
17	120	1	Urban	Υ	
125	360	1	Urban	Υ	
0	100	240		Urban	Υ

Árvore de Decisão

- Referencias
- Machine Learning. Tom Mitchell. McGraw-Hill.1997.
- https://www.cin.ufpe.br/~if684/EC/aulas/Aula-arvores-decisao-SI.p df

Foreman, John W. Data Smart: Using Data Science to Transform Information into Insight. Wiley. 1^a Ed., 2013.

- Curso IA para todos do prof. Diogo Cortiz da PUC-SP;
 https://www.youtube.com/watch?v=Ze-Q6ZNWpco&list=PLtQM10PgmGogjn0cikgWi8wpQUnV6ERkY
- Livros gratuitos de Ciência de Dadoshttps://www.dataquest.io/blog/data-science-books/
- Materiais do prof. Regis da UFC https://sites.google.com/site/ufcregis/home/data-science-machine-learning