

EDGARD JOSEPH KIRIYAMA

POSTECH

DATA ANALYTICS

MACHINE LEARNING COM PYTHON

AULA 02

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	11
O QUE VOCÊ VIU NESTA AULA?	12
REFERÊNCIAS	13
PALAVRAS-CHAVE	14

EMSE

O QUE VEM POR AÍ?

Olá, jovem analytic!

Você está na terceira disciplina do curso e aqui começa mais um ponto importantíssimo da jornada em que você está trilhando.

Até aqui você teve contato com a análise exploratória de dados, aprendeu a como manipular visualizações básicas de seus dados e uma boa base de estatística e álgebra linear.

Chegou o momento de aprimorarmos e elevarmos o nosso nível! Vamos te mostrar como utilizar os nossos dados e começar a aprofundar nestes conceitos sobre e suas aplicações de Machine Learning.

Vamos começar a entender como este aprendizado pode nos ajudar no nosso dia a dia! Bora lá?

CONHEÇA SOBRE O ASSUNTO

O Machine Learning (Aprendizado de Máquina) é um método de análise de dados que automatiza a construção de modelos analíticos.

Por ser um ramo da inteligência artificial, por estar baseado na ideia de que os sistemas podem aprender com dados, identificar padrões e tomar decisões com pouca intervenção humana, nós utilizamos algoritmos para a resolução de problemas e tomadas de decisão (SAS, 2023).

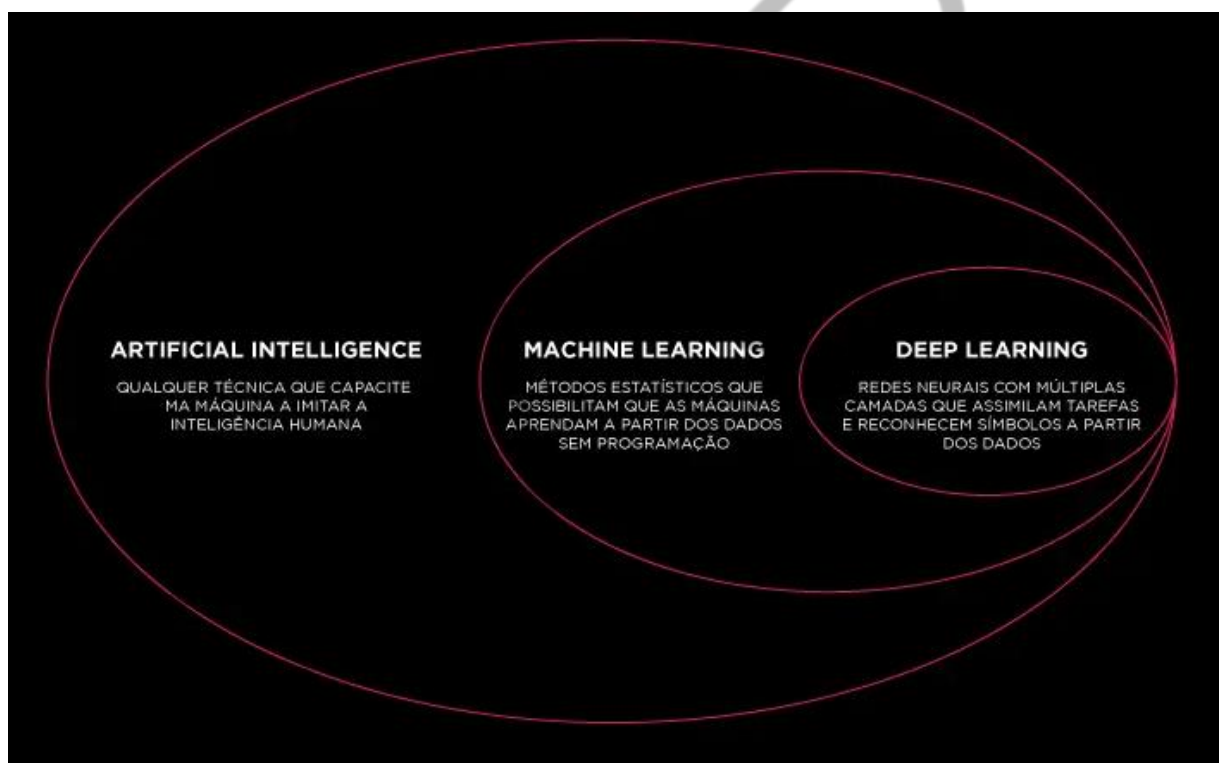


Figura 1 – Estrutura dos ramos da Inteligência Artificial

Fonte: <https://opencadd.com.br/machine-learning-ou-deep-learning/> (2019), adaptado por FIAP (2023)

A principal questão é: como desenvolver os modelos de Machine Learning (ML) e começar a desenvolver os nossos projetos? Talvez você esteja se perguntando: “Mas quais projetos?”. Vamos a alguns exemplos:

- Detecção de Fraudes em Cartões de Crédito;
- Mercado Financeiro;
- Vendas / Retenção de Clientes;
- Sistemas de Recomendação;

- Medicina;
- Telecomunicação.

Agora que já sabemos o que é Machine Learning, fica a pergunta: como utilizo esse aprendizado de máquina? Quais as técnicas?

Vamos começar pelos métodos clássicos:

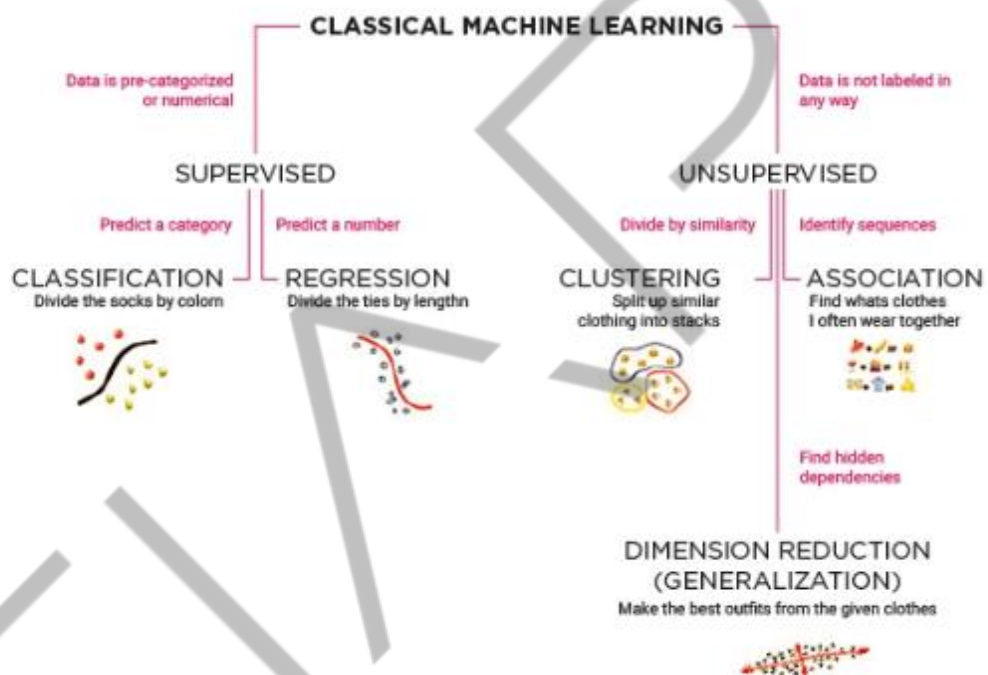


Figura 2 – Estrutura dos modelos clássicos de Machine Learning

Fonte: <https://www.datageeks.com.br/machine-learning/> (2019), adaptado por FIAP (2023)

- Aprendizagem Supervisionada
 - Usa um algoritmo que precisa de exemplos rotulados para desempenhar suas tarefas.
- Aprendizagem Não Supervisionada
 - Os algoritmos visam obter uma melhor representação dos dados, como em uma segmentação de clientes, ou recomendação de produtos que não usam rótulos pré-definidos.

Mas claro que Machine Learning não se resume somente a isso, conforme é possível visualizar na figura 3 – Estrutura dos modelos de Machine Learning

existentes. Mas não se preocupem! Vocês verão essas ferramentas no decorrer do curso.

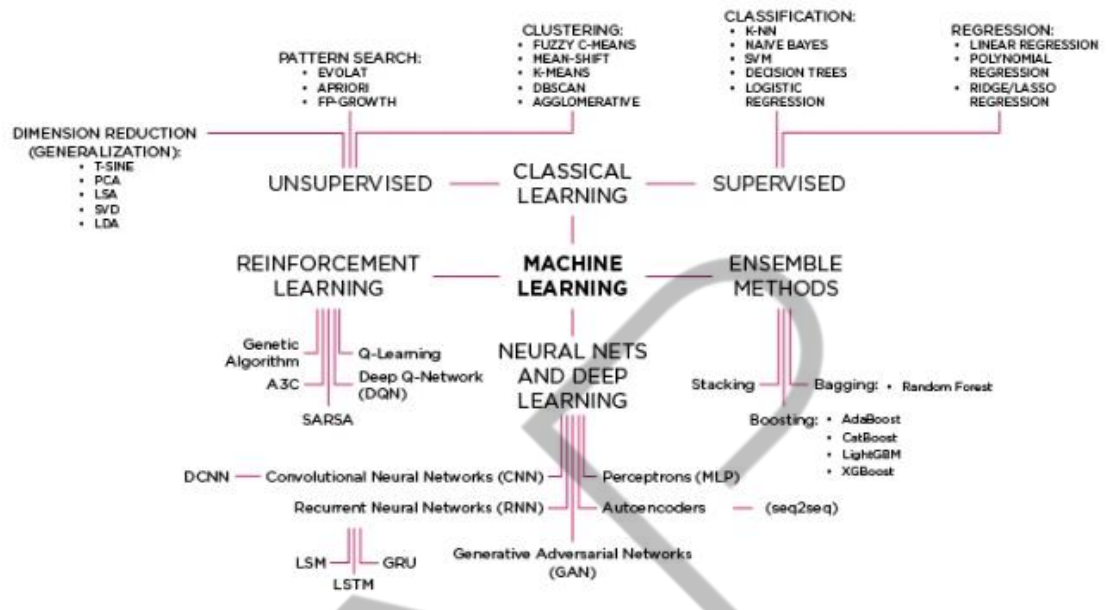


Figura 3 – Estrutura dos Modelos de Machine Learning existentes.

Fonte: <https://www.datageeks.com.br/machine-learning/> (2019), adaptado por FIAP (2023)

É muito importante entendermos que, para que tudo isso ocorra da melhor maneira possível, é necessário ocorrer todo um ciclo analítico (**pipeline**) para que nossos modelos funcionem corretamente, conforme podemos verificar na figura 4 – Ciclo analítico de Machine Learning.

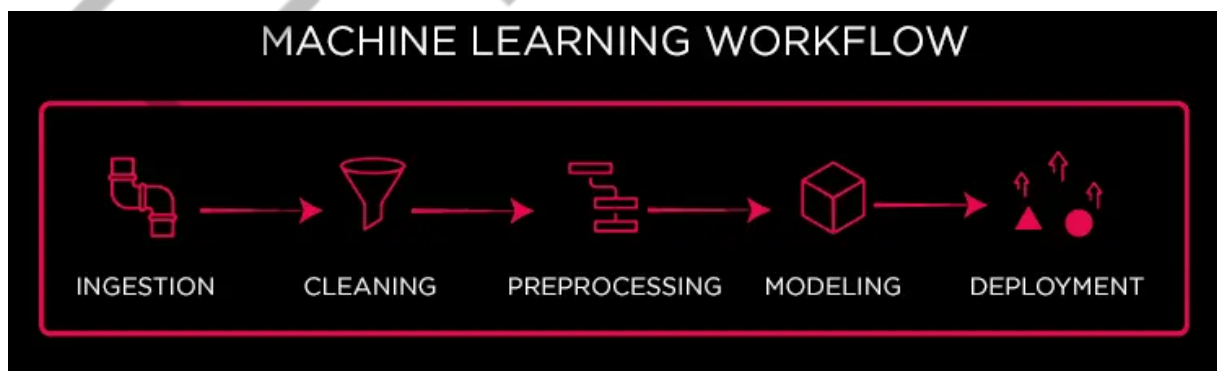


Figura 4 – Ciclo Analítico de Machine Learning.

Fonte: <https://www.javatpoint.com/machine-learning-pipeline> (s.d.), adaptado por FIAP (2023)

O ciclo analítico passa por diversos temas, que são:

- Entender o problema de Negócio (**etapa mais importante do ML**);

- Coletar os dados;
- Explorar / Visualizar;
- Feature Engineering (preparar os dados);
- Machine Learning;
- Validação / Monitoramento do Modelo.

Vamos passar pelas bibliotecas fundamentais que auxiliam no desenvolvimento dos MLs?

Já te aviso, as bibliotecas a seguir são as principais para o seu dia a dia, principalmente para quem está iniciando. É muito importante que você esteja ciente que existem inúmeras bibliotecas que nos ajudam na construção de modelos de ML. Há uma grande comunidade dedicada em desenvolver novos algoritmos para resolução de problemas. Veja na figura 5 - Principais algoritmos de Machine Learning como é extenso o estudo de Machine Learning.

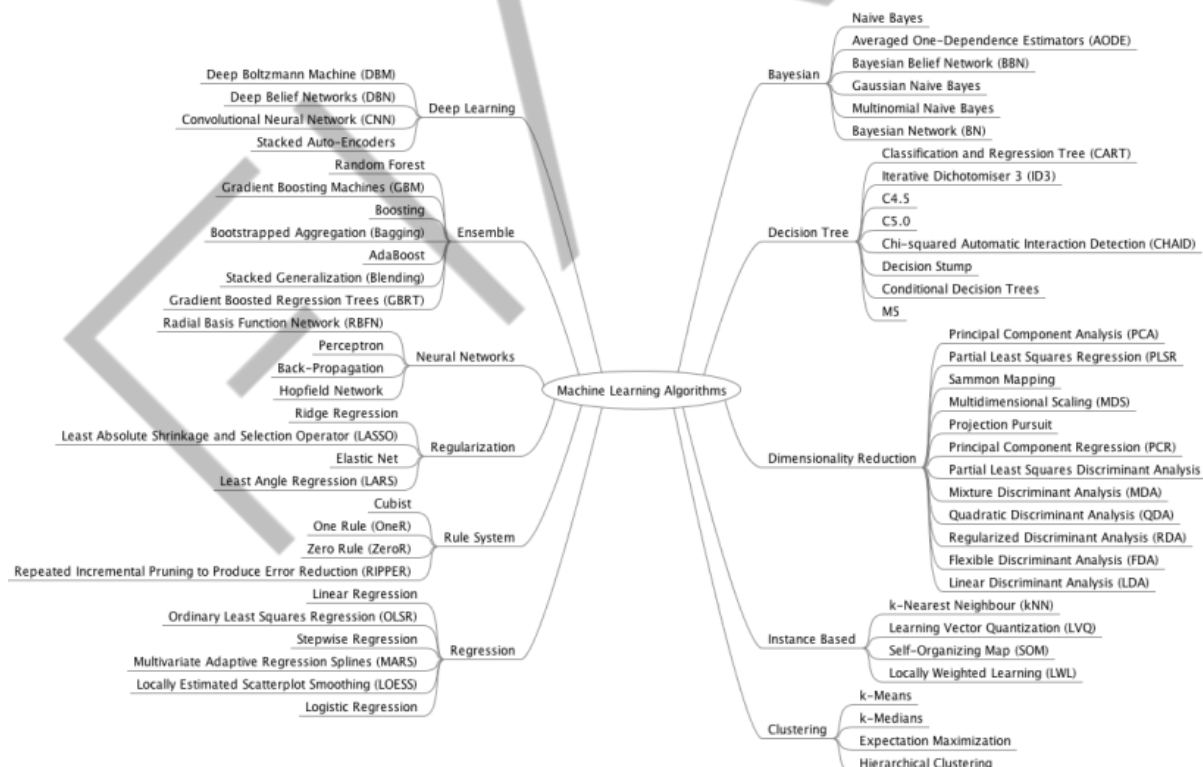


Figura 5 – Principais algoritmos de Machine Learning

Fonte: <https://jixta.wordpress.com/2015/07/17/machine-learning-algorithms-mindmap/> (2015)

As principais bibliotecas são:

NumPy

O Numpy é uma biblioteca do Python extremamente famosa por manipular matrizes multidimensionais, processamento de matrizes e por ter uma grande quantidade de funções matemáticas que auxiliam muito neste processo.

É uma ferramenta muito útil para os modelos de Machine Learning, pois possui uma grande variedade de rotinas para grandes operações de matrizes, lógica, manipulação de formas classificação, seleção, transformações discretas de Fourier, álgebra linear básica, operações estatísticas básicas, simulação aleatória entre outras.

A API NumPy auxilia em outras API famosas em Machine Learning, tais como Pandas, SciPy, Matplotlib e scikit-learn.

Quer obter mais informações? Consulte a [documentação do NumPy](#).

Pandas

Pandas é outra biblioteca muito popular para análise de dados. Provavelmente vocês já decoraram esse começo de código:

- `Import pandas as pd`

Essa API não está diretamente relacionada à construção de modelos de Machine Learning, porém é muito utilizada para a modelagem dos dados para que os modelos funcionem da melhor maneira possível.

Ela oferece uma estrutura de alto nível e com uma ampla variedade de ferramentas para a análise de dados antes e após rodarmos os modelos de ML.

Quer obter mais informações? Consulte a documentação abaixo:

[Documentação do Pandas](#).

Matplotlib

Outra biblioteca importante principalmente para a visualização de dados. Como vocês viram na disciplina passada, por ser uma biblioteca que tem o objetivo plotar

gráfico e figuras, ela não está diretamente relacionada às técnicas de construção de modelos de Machine Learning.

No entanto, ela nos ajuda a visualizar os resultados gerados a partir dos modelos, nos quais conseguimos avaliar se faz sentido o que está sendo processado ou não. Essa ferramenta também nos auxilia a visualizar o desempenho de nossos modelos através de gráficos.

Quer obter mais informações? Consulte a documentação abaixo:

[Documentação do Matplotlib.](#)

Scikit-Learn

Entrando a fundo em uma biblioteca típica de Machine Learning, o Scikit-Learn é a biblioteca mais popular, pois possui os principais algoritmos clássicos de ML.

Ela oferece modelos de algoritmos de aprendizagem supervisionados e não supervisionados. No entanto, essa biblioteca é utilizada também para mineração e análise de dados, se tornando assim uma ótima ferramenta de Machine Learning.

Quer obter mais informações? Consulte a documentação:

[Documentação do Scikit-Learn.](#)

TensorFlow

Outra biblioteca popular dentro dos algoritmos de Machine Learning, a TensorFlow. Ela foi desenvolvida pela equipe do Google com o objetivo de buscar um alto desempenho em cálculos numéricos através de uma estrutura tensores.

Muito utilizado para treinar e executar redes neurais profundas que podem ser usadas para desenvolver modelos de Inteligência Artificial.

Quer obter mais informações? Consulte a documentação:

[Documentação do TensorFlow.](#)

PyTorch

O PyTorch é uma biblioteca bastante utilizada para implementação de aprendizado profundo, como visão computacional e processamento de Linguagem

Natural (NLP). Essa biblioteca foi desenvolvida pela equipe de pesquisa de IA do Facebook (FAIR) em 2016. A biblioteca é ótima para cálculos matemáticos complexos por utilizar GPU. Um concorrente direto ao TensorFlow.

Quer obter mais informações? Consulte a documentação:

[Documentação do PyTorch.](#)

Dica de leitura:

Um bom jeito de entender melhor sobre Machine Learning é acompanhar as principais páginas de artigos e blogs da área.

Que tal dar uma lida [neste artigo](#) no blog do NEILPATEL? Aqui o autor traz uma ótima visão sobre Machine Learning e sua importância.

HANDS ON

Agora, chegou o momento de ver, na prática, como começar a importar nossos dados e trabalhar com eles via programação. A ideia é não se limitar apenas ao código explícito no hands on, então recomendamos que procure a documentação das bibliotecas, explore novas funcionalidades e muito mais!

EMANIP

O QUE VOCÊ VIU NESTA AULA?

Introdução ao conceito de Machine Learning e suas aplicações e começamos a ter contato com as principais bibliotecas fundamentais para começar a construção de um modelo.

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer mais suas bases e conhecimentos.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que assim você possa aprimorar os seus conhecimentos!

Você não está só nesta jornada! Te esperamos no Discord e nas lives com os nossos especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

DOCUMENTAÇÃO MATPLOTLIB. Disponível em: <https://matplotlib.org/stable/plot_types/index.html>. Acesso em: 04 mai 2023.

DOCUMENTAÇÃO NUMPY. Disponível em: <<https://numpy.org/numpy-tutorials/>>. Acesso em: 04 mai 2023.

DOCUMENTAÇÃO PANDAS. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 04 mai 2023.

DOCUMENTAÇÃO PYTORCH. Disponível em: <<https://pytorch.org/get-started/locally/>>. Acesso em: 04 mai 2023.

DOCUMENTAÇÃO SCIKIT-LEARN. Disponível em: <<https://scikit-learn.org/stable/modules/classes.html>>. Acesso em: 04 mai 2023.

DOCUMENTAÇÃO TENSORFLOW. Disponível em: <https://www.tensorflow.org/api_docs/python/tf>. Acesso em: 04 mai 2023.

GRUS, Joel. **Data Science do Zero**. Rio de Janeiro: Alta Books Editora, 2016.

MCKINNEY, Wes. **Python para análise de dados**: tratamento de dados com pandas, numpy e ipython. São Paulo: Novatec Editora, 2019.

SAS. Disponível em: <https://www.sas.com/pt_br/insights/analytics/machine-learning.html>. Acesso em: 04 mai 2023.

PALAVRAS-CHAVE

Palavras-chave: Python. Pandas. Machine Learning.

EMENDAS

The background is a dark blue field filled with numerous small, light blue dots, resembling a starry sky or a data visualization. Overlaid on this are several large, wavy, translucent lines in shades of blue, teal, and yellow, which flow across the page. Scattered throughout are various geometric shapes: a circle containing the number '7' in the upper center, a small circle on the left, a cross-like shape near the bottom left, a small circle near the bottom center, and a hexagon in the bottom right corner.

POSTECH