The background of the slide features a close-up of a man's face, partially obscured by a large, semi-transparent teal triangle on the left. Overlaid on the image are various digital and network-themed graphics, including glowing blue lines, nodes, and hexagons. In the upper right, there are some numerical data points like '51.07' and '51.84'.

Analytics e Inteligência Artificial Data Science

Tema da aula
Introdução ao Machine Learning



BUSINESS SCHOOL

Graduação, pós-graduação,
MBA, Pós- MBA, Mestrado
Profissional, Curso In
Company e EAD



CONSULTING

Consultoria personalizada
que oferece soluções
baseadas em seu
problema de negócio



RESEARCH

Atualização dos
conhecimentos e do material
didático oferecidos nas
atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais **escolas de negócio do mundo**, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 **projetos de consultorias** em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única *Business School* brasileira a figurar no *ranking* LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação



O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de **Big Data, Analytics** e **Inteligência Artificial**.



Profª Drª Alessandra Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil. Os diretores foram professores de grandes especialistas do mercado.

- +10 anos de atuação.
- +9.000 alunos formados.

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria;
- Larga experiência de mercado na resolução de *cases*;
- Participação em congressos nacionais e internacionais;
- Professor assistente que acompanha o aluno durante todo o curso.

Estrutura

- 100% das aulas realizadas em laboratórios;
- Computadores para uso individual durante as aulas;
- 5 laboratórios de alta qualidade (investimento +R\$2MM);
- 2 unidades próximas à estação de metrô (com estacionamento).



PROFA. DRA. ALESSANDRA DE ÁVILA MONTINI

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Tem muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e Inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em Estatística Aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Parecerista da FAPESP e colunista de grandes portais de tecnologia.





PROF. ÂNGELO CHIODE, MSc

Bacharel, mestre e candidato ao PhD em Estatística (IME-USP), atua como professor de Estatística Aplicada para turmas de especialização, pós-graduação e MBA na FIA. Trabalha como consultor nas áreas de Analytics e Ciência de Dados há 13 anos, apoiando empresas na resolução de desafios de negócio nos contextos de finanças, aquisição, seguros, varejo, tecnologia, aviação, telecomunicações, entretenimento e saúde. Nos últimos 5 anos, tem atuado na gestão corporativa de times de Analytics, conduzindo projetos que envolviam análise estatística, modelagem preditiva e *machine learning*. É especializado em técnicas de visualização de dados e design da informação (Harvard) e foi indicado ao prêmio de Profissional do Ano na categoria Business Intelligence, em 2019, pela Associação Brasileira de Agentes Digitais (ABRADi).



Conteúdo Programático

6



DISCIPLINAS



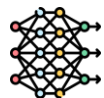
**IA E TRANSFORMAÇÃO
DIGITAL**



ANALYTICS



**INTELIGÊNCIA ARTIFICIAL:
MACHINE LEARNING**



**INTELIGÊNCIA ARTIFICIAL:
DEEP LEARNING**



**EMPREENDEDORISMO E
INOVAÇÃO**



**COMPORTAMENTO
HUMANO E SOFT SKILLS**

TEMAS: ANALYTICS E MACHINE LEARNING

ANÁLISE EXPLORATÓRIA DE DADOS

INFERÊNCIA ESTATÍSTICA

TÉCNICAS DE PROJEÇÃO

TÉCNICAS DE CLASSIFICAÇÃO

TÓPICOS DE MODELAGEM

TÉCNICAS DE SEGMENTAÇÃO

TÓPICOS DE ANALYTICS

MANIPULAÇÃO DE BASE DE DADOS

AUTO ML

TEMAS: DEEP LEARNING

REDES DENSAS

REDES CONVOLUCIONAIS

REDES RECORRENTES

MODELOS GENERATIVOS

FERRAMENTAS

LINGUAGEM R

LINGUAGEM PYTHON

DATABRICKS



Conteúdo da Aula

- 1. Introdução
- 2. Tópicos Abordados
- Referências Bibliográficas



1. Introdução

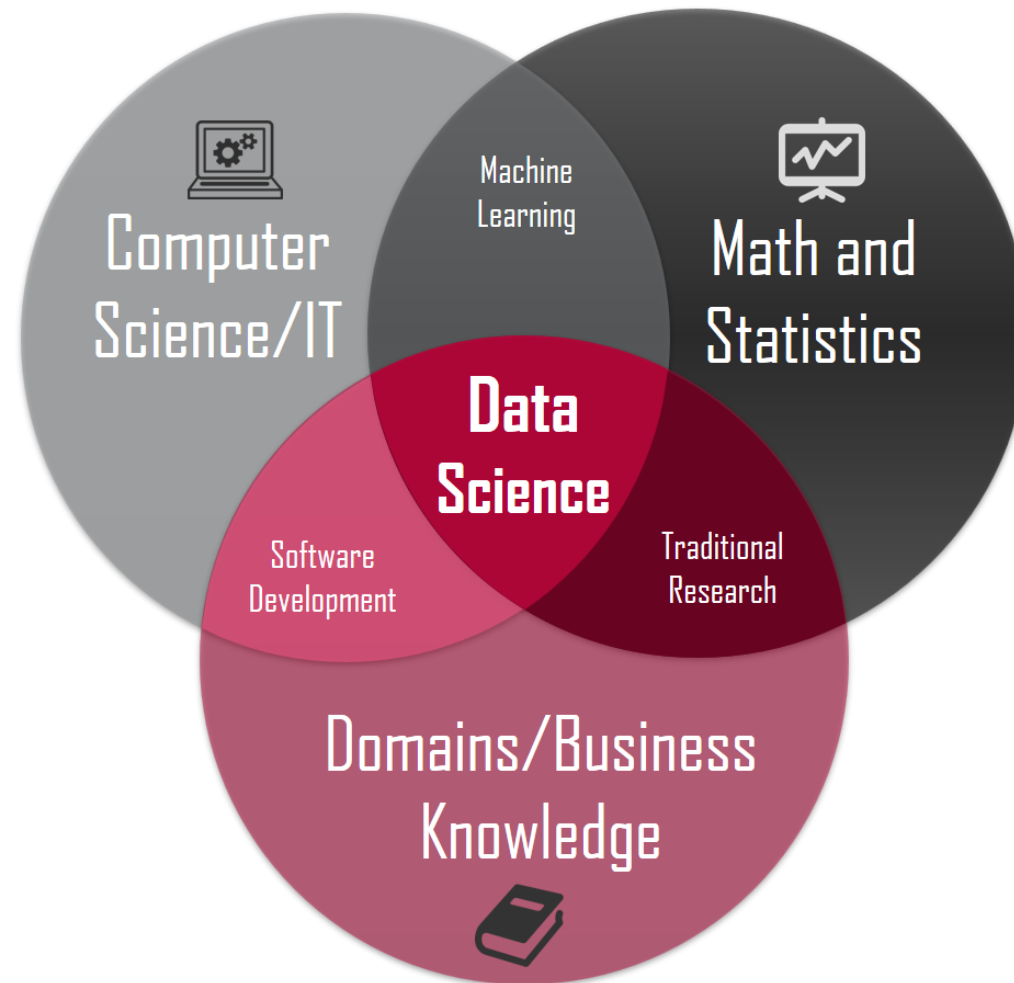




Definição de *Machine Learning*

1. INTRODUÇÃO | INTRODUÇÃO AO MACHINE LEARNING

9



Créditos da imagem: <https://medium.com/@anuraggandhi29/what-is-datascience-6ac639f830c2>



Definição de *Machine Learning*

1. INTRODUÇÃO | INTRODUÇÃO AO MACHINE LEARNING

10

O **aprendizado de máquina**, ou *machine learning*, engloba o desenvolvimento de algoritmos que permitem que computadores identifiquem padrões e realizem previsões a partir de dados históricos, sem serem explicitamente programados para cada nova tarefa.

A definição do termo foi realizada por Arthur Samuel durante seus trabalhos na IBM, em 1959.

Exemplos de aplicações:

- **Detecção de fraudes:** Identificação de atividades suspeitas e prevenção de fraudes em transações financeiras, especialmente em bancos e empresas de cartões de crédito.
- **Risco de crédito:** Avaliação da capacidade de pagamento de clientes com base em seu histórico no mercado, para decidir sobre a concessão de crédito em instituições financeiras.
- **Churn de clientes:** Identificação de indivíduos com alto potencial de deixar de utilizar um serviço ou produto, permitindo a tomada de decisão antecipada para retenção de clientes.



Statistical Learning (SL)

No módulo de Analytics do nosso curso, havíamos nos concentrado nesta abordagem, que preza por:

- Avaliar a adesão dos dados às **premissas teóricas** dos modelos.
- Compreender a **relação** entre as variáveis.
- **Interpretar** os resultados obtidos.

Machine/Deep Learning (ML/DL)

A partir de agora, nossos principais focos são:

- Obter algoritmos com **alta qualidade de predição**, ou seja, que consigam realizar boas predições ou classificações, sem necessariamente respeitarem suposições ou serem de fácil interpretação.
- Desenvolver algoritmos **mais complexos**, voltados para identificar fenômenos também complexos, lançando mão de infraestrutura computacional robusta (*big data*).



Statistical Learning vs. Machine Learning

1. INTRODUÇÃO | INTRODUÇÃO AO MACHINE LEARNING

12

Observação

Os modelos estatísticos que estudamos até então também podem ser considerados modelos de *machine learning*, pois o foco deles também é o **aprendizado** baseado em dados para fins de predição ou segmentação. O que muda é apenas a ótica/objetivo: maior ênfase na compreensão estrutural do fenômeno e na interpretação (SL) ou apenas na qualidade da predição/classificação em si (ML).



2. Tópicos Abordados





Tópicos Abordados

2. TÓPICOS ABORDADOS | INTRODUÇÃO AO MACHINE LEARNING

14

Vamos nos aprofundar nos seguintes tópicos em *machine learning*:

1. Linguagem Python
2. Framework geral de modelagem
3. Manipulação de dados para modelagem
4. Técnicas/algoritmos de classificação
5. Técnicas/algoritmos de projeção
6. Técnicas/algoritmos de segmentação
7. *Automatic machine learning* (Auto ML)
8. MLOps





Generalidade: O Python é uma linguagem de programação de propósito geral, o que significa que ela pode ser utilizada para uma variedade de tarefas além do campo da ciência de dados, tais como desenvolvimento web, automação, entre outros.

Machine Learning: O Python é considerado mais forte quando se trata de áreas como *machine learning* e inteligência artificial, pois possui pacotes com maior quantidade de recursos e maior popularidade, como *TensorFlow*, *PyTorch* e *Scikit-Learn*.



Framework Geral de Modelagem

2. TÓPICOS ABORDADOS | INTRODUÇÃO AO MACHINE LEARNING

16

Neste tema, estudaremos o **passo-a-passo geral** envolvido na construção de um modelo ou algoritmo, com ênfase para tópicos relacionados a *machine learning*.



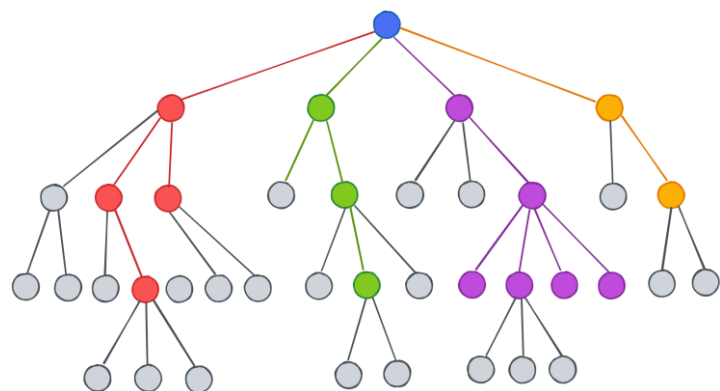
Algumas boas práticas de **engenharia de variáveis** (*feature engineering*) podem ser empregadas para fins de preparação de dados e apoio na melhoria de qualidade de um futuro algoritmo.

Podemos destacar os seguintes métodos:

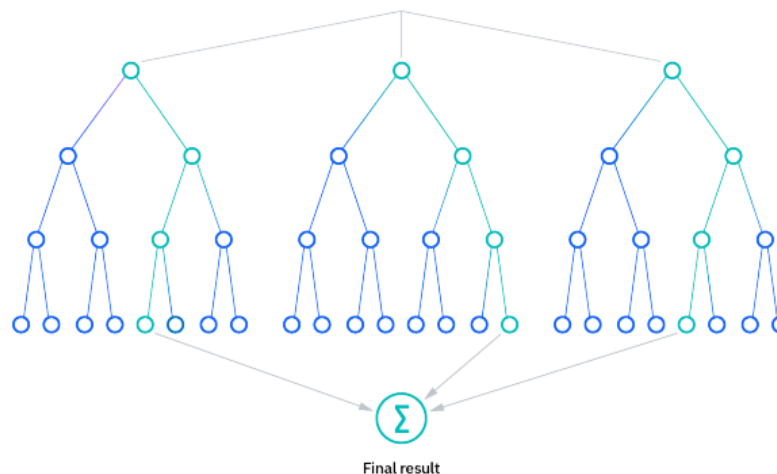
- **Criação de variáveis** (*feature construction*), especialmente quando há **escassez** de variáveis originais.
- **Redução de dimensionalidade** (*dimensionality reduction*), especialmente quando há **excesso** de variáveis originais.



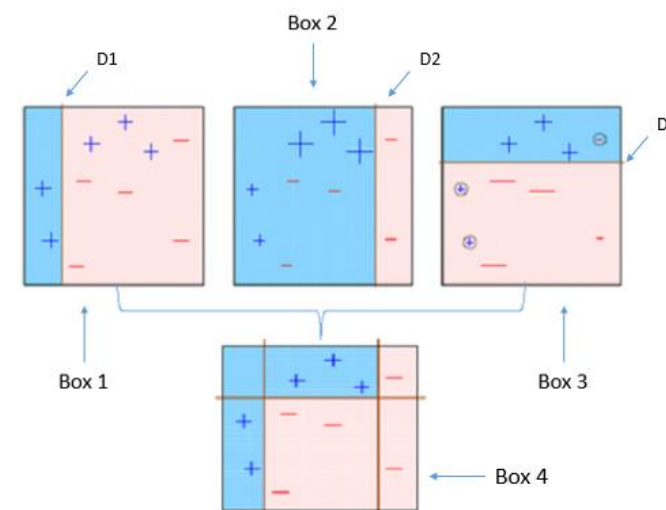
Extensão do **modelo de regressão logística** para algoritmos computacionais com objetivo de prever a **categoria** (*target*) a qual um objeto pertence, em função de outros aspectos **explicativos** (*features*).



Árvore de Decisão



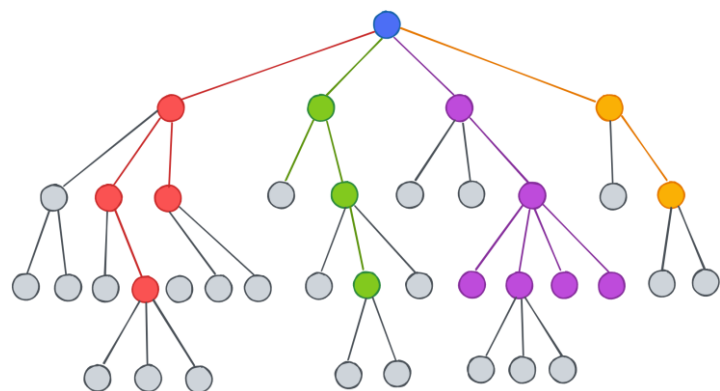
Random Forest



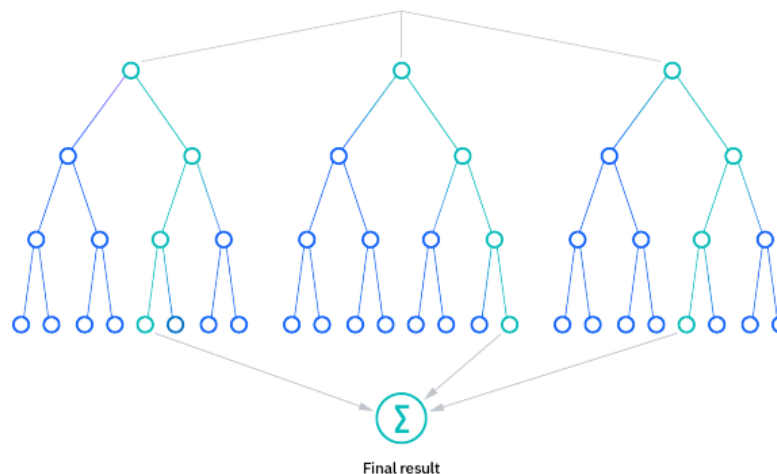
Gradient Boosting
(diversas variantes)



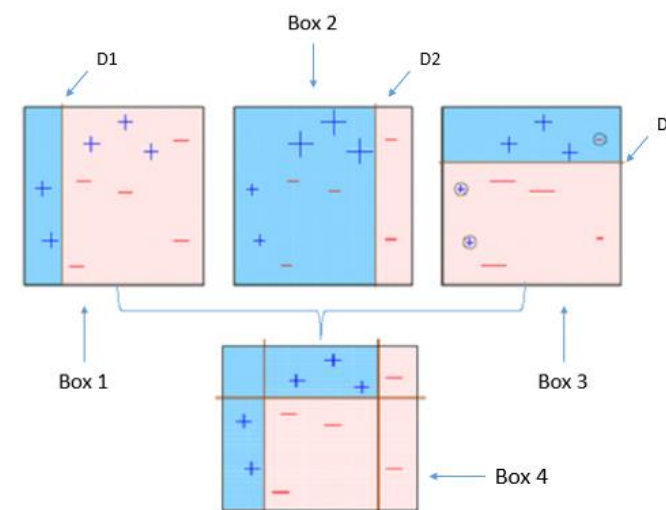
Extensão do **modelo de regressão linear** para algoritmos computacionais com objetivo de projetar/predizer o valor de um **aspecto quantitativo de interesse** (*target*) em função de outros aspectos **explicativos** (*features*).



Árvore de Regressão



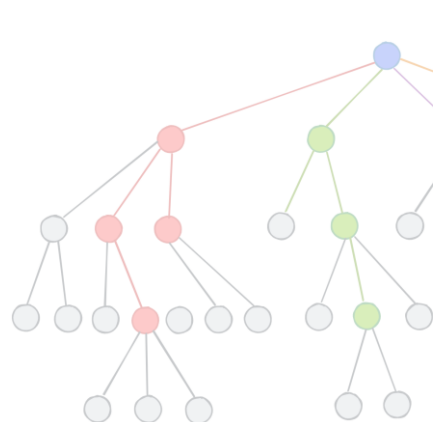
Random Forest



Gradient Boosting
(diversas variantes)



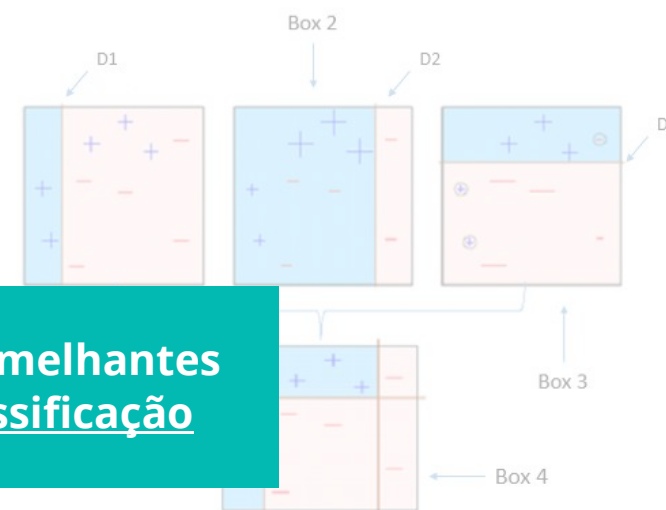
Extensão do **modelo de regressão linear** para algoritmos computacionais com objetivo de projetar/predizer o valor de um **aspecto quantitativo de interesse** (*target*) em função de outros aspectos **explicativos** (*features*).



As técnicas mais utilizadas para dados tabulares são semelhantes para problemas de regressão e para problemas de classificação



Final result



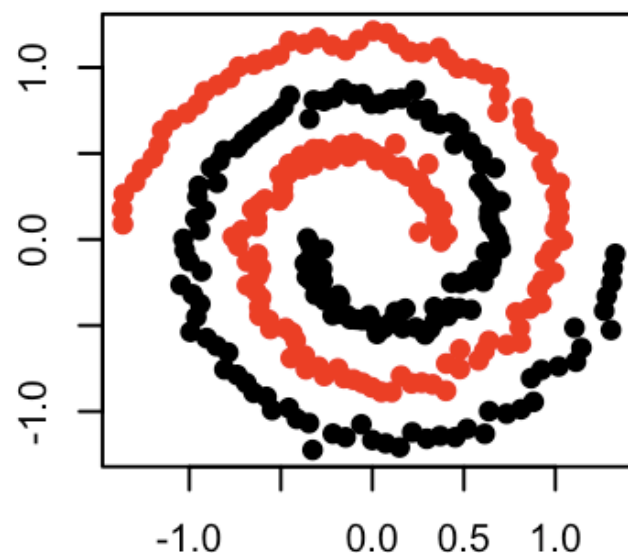
Árvore de Regressão

Random Forest

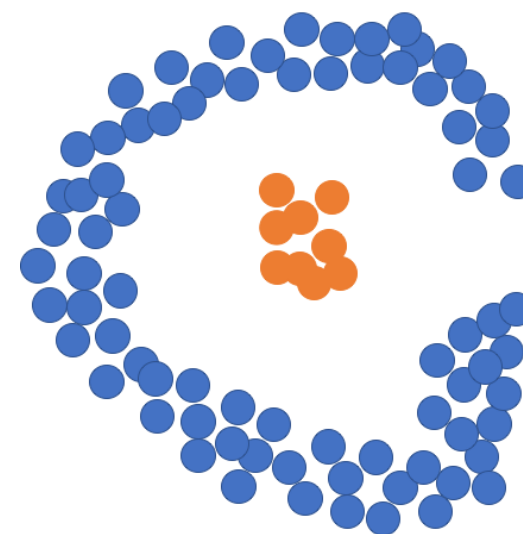
Gradient Boosting
(diversas variantes)



Extensão dos algoritmos tradicionais de **análise de cluster** para outros algoritmos computacionais mais sofisticados, com objetivo de **agrupar elementos** com base em sua **semelhança**.



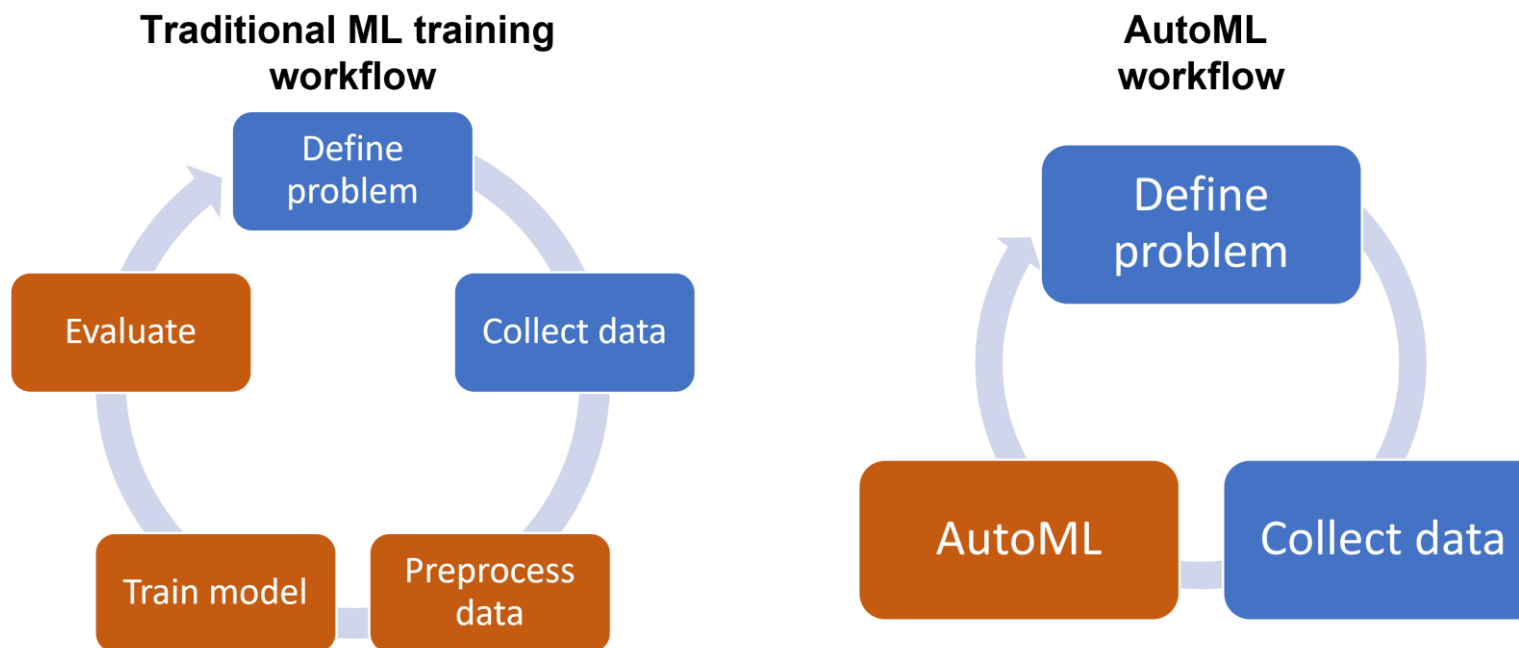
Clusterização
Espectral



Clusterização Baseada
em Densidade (DBSCAN)



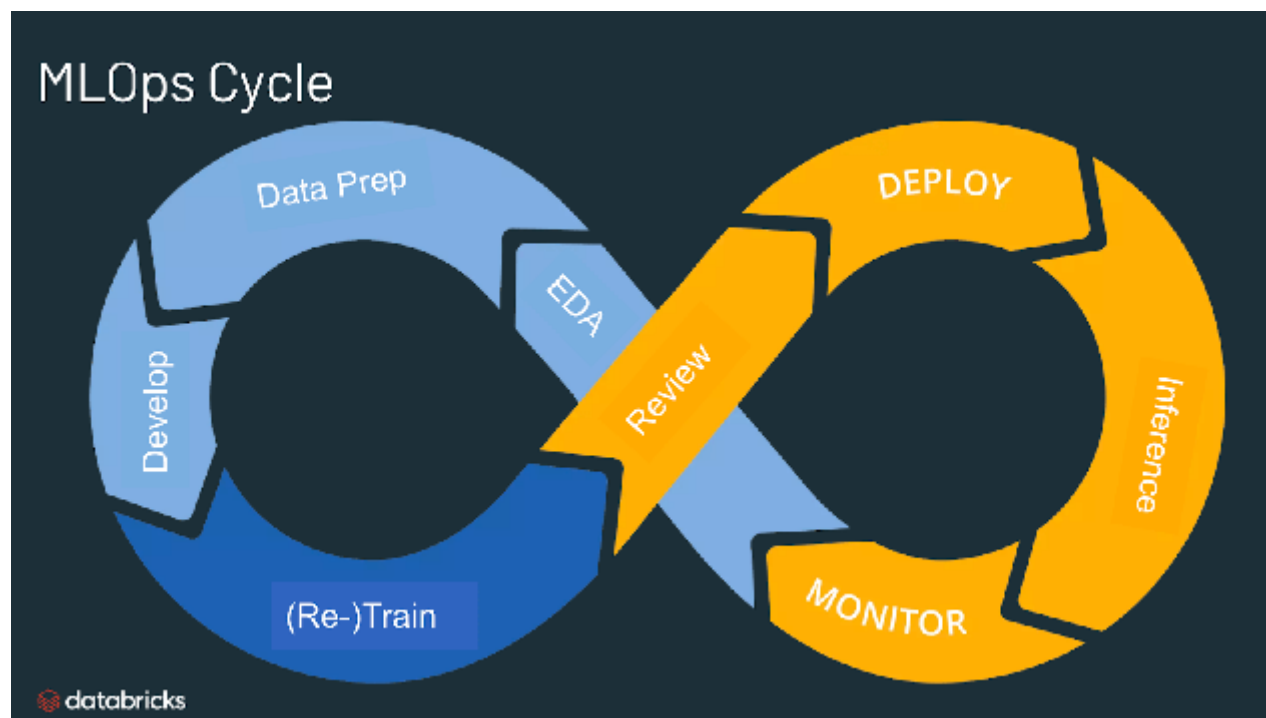
Cuidados para **automatização** do desenvolvimento de **algoritmos supervisionados**, o que inclui processamento de dados, seleção de algoritmos e otimização de hiperparâmetros, visando a manutenção de performance.



Créditos da imagem: <https://learn.microsoft.com/pt-br/dotnet/machine-learning/automated-machine-learning-mlnet>



O **MLOps** (*Machine Learning Operations*) abrange todo o ciclo de vida de algoritmos de ML: desenvolvimento, testes, monitoramento, recalibrações, *deploy* (implantação em ambiente produtivo) e consumo estável dos usuários finais.



Créditos da imagem: <https://www.databricks.com/glossary/mlops>



Referências Bibliográficas

INTRODUÇÃO AO MACHINE LEARNING

24

- James, G. *An Introduction to Statistical Learning - With Applications in R*. 2ª edição. Springer, 2021.
- Kubat, M. *An Introduction to Machine Learning*. 2ª edição, Springer, 2017.
- Pinheiro, C. A. R, Patetta, M. *Introduction to Statistical and Machine Learning Methods for Data Science*. SAS, 2021.





lab.data

<http://labdata.fia.com.br>
Instagram: @labdatafia
Facebook: @LabdataFIA

