

## Fecomércio Sesc

Data Science – Princípios e Técnicas

Maio

2025



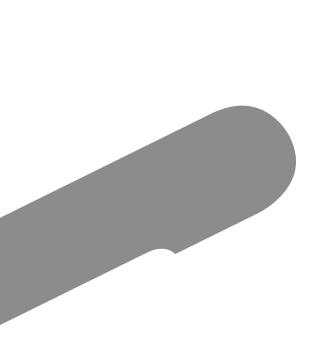
Onde me encontrar:

https://www.linkedin.com/in/marco-mialaret-junior/

e

https://github.com/MatmJr









O aprendizado de máquina é uma área da inteligência artificial que permite que computadores aprendam a realizar tarefas a partir de dados, sem a necessidade de programação explícita para cada situação específica.



Ao invés de seguir instruções fixas, os algoritmos de machine learning analisam grandes volumes de informações, reconhecem padrões e fazem previsões ou decisões com base nesses padrões.



O objetivo central do aprendizado de máquina é construir modelos que generalizem bem, ou seja, que consigam produzir boas respostas mesmo quando expostos a dados que não foram vistos durante o treinamento.



Para isso, os modelos são treinados com um conjunto de dados históricos e depois testados em novos dados para avaliar sua capacidade de previsão.



Essa abordagem tem ganhado espaço em diversas áreas, como diagnóstico médico, recomendação de produtos, previsão de demanda, reconhecimento de voz e imagem, entre muitas outras.





### Algoritmos Supervisionados e Não Supervisionados



Os algoritmos de aprendizado de máquina são tradicionalmente classificados em duas grandes categorias: supervisionados e não supervisionados, dependendo do tipo de dado e do objetivo da tarefa.



Nos algoritmos supervisionados, o modelo é treinado com dados rotulados, ou seja, cada exemplo de entrada está associado a uma saída conhecida. O modelo aprende essa relação entre entrada e saída para conseguir prever resultados futuros.



Essa abordagem é usada, por exemplo, para prever preços de imóveis com base em características como tamanho, localização e número de quartos (regressão), ou para identificar se um e-mail é spam ou não (classificação).



Já os algoritmos **não supervisionados** trabalham com dados que não têm rótulos definidos. Nesse caso, o objetivo é explorar a estrutura dos dados, buscando padrões ou agrupamentos naturais.



Um exemplo clássico é a segmentação de clientes, em que o modelo identifica grupos com comportamentos semelhantes a partir de informações como histórico de compras, idade ou localização, sem que esses grupos estejam previamente definidos.



Além dessas duas categorias, há também abordagens como o aprendizado por reforço, que é mais utilizado em sistemas de decisão autônoma, como robôs ou jogos, onde o agente aprende com base em recompensas e penalidades por suas ações.





# Nosso primeiro algoritmo de ML



Dentro dos algoritmos supervisionados, um dos mais populares é o Random Forest, ou floresta aleatória. Ele pertence à família dos métodos baseados em árvores de decisão.



Uma árvore de decisão é uma estrutura que simula um processo de tomada de decisão em que cada nó representa uma pergunta ou condição sobre os dados, e cada ramo representa o resultado dessa decisão, levando a um novo nó ou a uma resposta final.



Embora as árvores de decisão sejam fáceis de interpretar, elas podem ser instáveis e muito sensíveis a variações nos dados. Para superar esse problema, o Random Forest combina várias árvores de decisão em um único modelo.



Cada árvore da floresta é treinada com uma amostra aleatória dos dados e com uma seleção aleatória de variáveis. Ao diversificar os dados e as variáveis, o modelo evita o risco de que todas as árvores cometam os mesmos erros.



No momento da previsão, cada árvore gera uma resposta individual. No caso de problemas de classificação, a resposta final será a classe mais votada entre todas as árvores.



Em problemas de regressão, a resposta final é a média das previsões de todas as árvores. Essa abordagem coletiva torna o Random Forest um modelo muito robusto, capaz de lidar com dados ruidosos, variáveis correlacionadas e uma grande quantidade de informações.



Slide:

https://drive.google.com/file/d/1K3zBTG3DIaVSfT3sft aalUcjISd9JHZ4/view?usp=drive\_link



### **Dúvidas?**





#### **Marco Mialaret, MSc**

**Telefone:** 

81 98160 7018

E-mail:

marco.junior@pe.senac.br

