

Disciplina:

# Machine Learning

Professor: Gabriel Assunção



## Quem sou eu?

Gabriel Oliveira Assunção

Graduação em Estatística UFMG;

Mestrado em Estatística UFMG;

Doutorando em Estatística UFMG

Cientista de Dados na Localiza&Co



[/gabrieloassuncao](https://www.linkedin.com/in/gabrieloassuncao)

# Sobre vocês

# **Apresentação do curso**

## **Regras gerais**

- Teremos duas pausas durante as aulas.
- Comunicação inteiramente pelo Teams
- Aulas às terças das 19:00 às 22:30
- Tarefas disponibilizadas nos canvas juntamente com os materiais

# O que é aprendizado de máquina?

# Apresentação do curso

## Objetivo

- Processo de aprendizagem de máquina.
- Tipos de dados.
- Técnicas e algoritmos de aprendizado supervisionado e não-supervisionado.
- Combinação de modelos.
- Métricas e avaliação de modelos.

# Apresentação do curso

## Objetivo

Ao final do curso iremos conseguir:

- Definir **qual a melhor abordagem para resolução do nosso problema**. Por exemplo: queremos classificar um indivíduo como possível cliente que irá comprar.
- Utilizar de **diferentes técnicas** para solucionar o nosso problema.
- Definir qual a **melhor técnica**
- **Interpretar os resultados** do nosso modelo

# Apresentação do curso

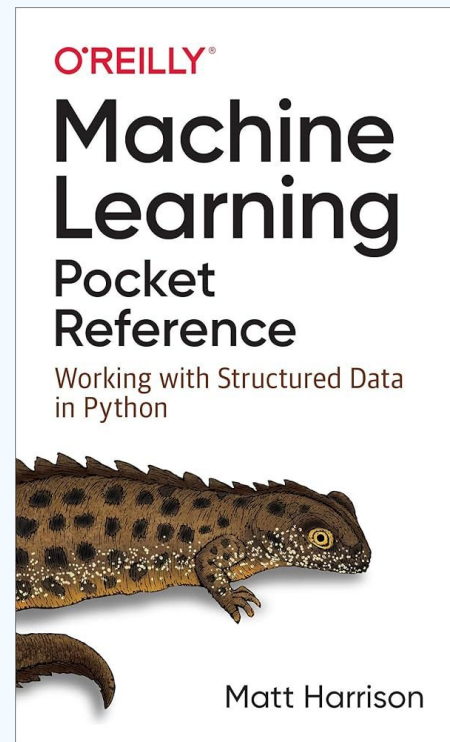
## Módulos

1. Introdução
2. Modelo supervisionado
  - a. Problemas de regressão, exemplo prever o valor de um produto
  - b. Problemas de classificação, exemplo prever a probabilidade de um cliente comprar.
3. Combinação de modelos
4. Modelos não supervisionados
  - a. Técnicas de visualização de dados
  - b. Agrupamento, exemplo criar segmentação de clientes
5. Tópicos avançados



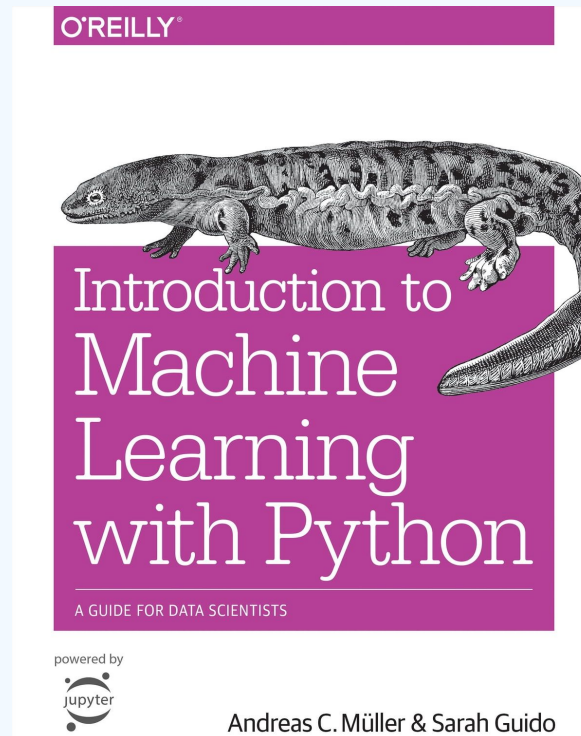
# Apresentação do curso

## Referências



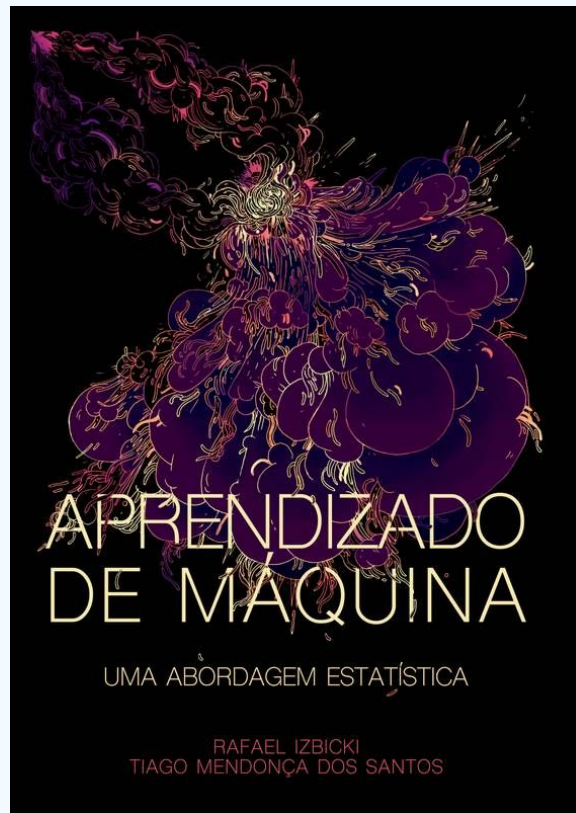
# Apresentação do curso

## Referências



# Apresentação do curso

## Recomendação



# **Apresentação do curso**

## **Avaliação**

1. Três listas no valor de 20 pontos cada (em grupo de até 4 pessoas)
2. Um trabalho prático no valor de 40 pontos (em grupo de 2 até 4 pessoas):
  - a. Apresentação do dados e do problema (5 pontos)
  - b. Código para aplicação (25 pontos)
  - c. Relatório com análises, modelos e conclusões (10 pontos)

# Apresentação do curso

## Listas

1. Três listas sobre os módulos 2, 3 e 4.
2. Cada lista irá valer 20 pontos
3. A cada dia de atraso na entrega da lista será descontado 2 pontos da nota.

# Apresentação do curso

## Trabalho prático

- Cada grupo deverá selecionar um banco de dados onde se é possível fazer uma aplicação de machine learning. Uma fonte para a busca de banco é o site kaggle.
- O grupo deve entregar um relatório contendo uma breve descrição do banco (fonte do dado, variáveis do dado e o objetivo da análise)
- Para a aplicação será avaliada a análise exploratória dos dados, pré-processamento, treino de modelos e avaliação de desempenho.
- Ao final deve ser entregue o código utilizado e um relatório (pode estar junto com o código) contendo:
  - a descrição da análise exploratória
  - as etapas do pré-processamento (com a justificativa de cada uma)
  - modelos testados
  - métrica utilizada para a escolha do modelo
  - conclusão de qual modelo selecionado.