



WBA0753_v1.0

Linguagens de programação para ciência de dados (*Python com Spark*)



Machine Learning em Python

Introdução. Modelos supervisionados e não supervisionados. Exemplos práticos.

Bloco 1

Yuri Sá



➤ Introdução ao aprendizado de máquina

Resumidamente, podemos dizer que aprendizado de máquina é o conceito de ensinar máquinas (computadores) a executar tarefas com ou sem supervisão, por meio de algoritmos de programação (MÜLLER, 2016; MARSLAND, 2014).



► Grandes tipos de modelos de *Machine Learning*

- Modelos supervisionados:
 - Classificação.
 - Regressão.
- Não supervisionado:
 - Clusterização.



➤ Módulos *Python*

Módulos utilizados:

- *NumPy.*
- *scikit-learn.*
- *SciPy.*
- *Matplotlib.*
- *Pandas.*



Machine Learning em Python

Introdução. Modelos supervisionados e não supervisionados. Exemplos práticos.

Bloco 2

Yuri Sá





Regressão linear

- Boston *house prices dataset*.
- *Split data*.
- Modelo: *LinearRegression*.
- Previsão.
- Métrica.



► Classificação

- *Breast Cancer Wisconsin Diagnostic Database.*
- *Split data (70/30).*
- Modelo: Naive Bayes.
- Previsão.
- Métrica.



► Clusterização

- *make_blobs.*
- *Plot.*
- Modelo: *k-means.*
- *Plot dos clusters.*



Teoria em Prática

Bloco 3

Yuri Sá



Reflita sobre a seguinte situação

- Produção de soja prejudicada por pragas.
- Identificar regiões mais afetadas.
- Separar as imagens, em grupos ou categorias, pelos tipos de pragas.



➤ Norte para a resolução...

- Identifique as bibliotecas para lidar com as imagens.
- Imagens são tabelas de valores de pixels.
- Estude o problema para definir o tipo de modelo.
- Defina a melhor estratégia para a classificação das imagens.



Dica do (a) Professor (a)

Bloco 4

Yuri Sá





Escolha do modelo

- Estude as características de cada modelo, suas fraquezas e potenciais.
- Conheça os dados e suas especificidades, muito tempo e recursos podem ser economizados só limpando bem os dados (cuidado para não perder informação).
- Não se prenda a um só modelo, teste outros, nunca fique preso à receitas tradicionais. Os modelos de *Machine Learning* são muito nsensíveis ao *dataset* aplicado e os resultados podem diferir muito da receita.





Referências

MARSLAND, S. **Machine learning**: an algorithmic perspective. Chapman and Hall/CRC, 2014.

MÜLLER, A. C.; SARAH, G. Introduction to machine learning with Python: a guide for data scientists. **O'Reilly Media**, Inc., 2016



Bons estudos!

