

Python para Ciência de Dados

Trabalho Prático

Prof. Rodrigo Silva

1 Leitura Recomendada

- VanderPlas, Jake. Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc., 2016. Disponível em <https://jakevdp.github.io/PythonDataScienceHandbook/>

2 Atividade Prática

1. Apresentar e resolver algum aplicação de ciência de dados com as seguintes características.
 - (a) A aplicação deve conter um problema de regressão ou de classificação.
 - (b) A aplicação deve envolver pelo menos duas fontes de dados.

3 Entregável

1. O aluno deverá produzir um notebook python que apresente com os seguintes requisitos mínimos:
 - (a) Carregamento das fontes de dados.
 - (b) Fusão das fontes de dados (`merge` e/ou `join`)
 - (c) Análise descritiva dos dados (Média, Max, Min)
 - (d) Análise gráfica dos dados
 - (e) Treinamento, ajuste fino e teste de modelos de aprendizado de máquina.

Cada uma destas etapas deve ser comentada (no próprio notebook com markdown) com uma descrição do que levou à escolha das técnicas e quais as conclusões das etapas.

4 Entrega

- O trabalho deve ser **finalizado até o dia 18/06**.
- Os alunos devem encaminhar pelo moodle apenas o link para o notebook que deve estar disponível e público em um repositório no GitHub.