

Python para Ciência de Dados

Pré-requisitos:

- Python

Carga horária: 32 horas

Texto de chamada: Ciência de dados é um dos assuntos mais importantes na atualidade. Com crescentes bases de dados à disposição é comum nos perguntarmos quais informações valiosas podem estar presentes ali mas inacessíveis a olho nu. A linguagem Python é uma das principais escolhas atualmente quando é necessário analisar, tratar, filtrar ou fazer qualquer tipo de manipulação de dados em massa. Possui ferramentas poderosas como NumPy, Pandas, Scikit-learn, Matplotlib dentre outras que auxiliam no trabalho de um cientista de dados. Todos os conceitos de ciência de dados estão muito relacionados ao também muito comentado atualmente, aprendizado de máquina. Uma introdução sobre este assunto também é abordada neste curso.

Conteúdo Programático:

- Revisão de Python 3
- Configuração do ambiente
- IPython e Jupyter
 - Notebooks
 - Comandos mágicos
 - Uso pelo terminal e pelo navegador
 - Comandos
- Revisão de estatística
 - Variáveis aleatórias
 - Distribuições de Probabilidade
 - Histograma
 - Média, mediana
 - Desvio padrão e variância populacionais e amostrais
 - Quartis
 - Boxplots
- NumPy
 - Arrays e dimensões
 - Indexing e Slicing
 - Boolean Indexing
 - Funções universais
 - Desempenho
 - Vetorização
 - Broadcasting
 - Métodos matemáticos e estatísticos
 - Operações orientadas a arrays
 - Álgebra Linear
- Pandas
 - Series

- DataFrame
- Indexação, Reindexação, Seleção e Filtragem
- Múltiplos Índices
- loc e iloc
- Concatenação e Joins
- Agregação e Agrupamentos
- Pivot Tables
- Trabalhando com séries temporais
- Carregamento de dados
 - CSV
 - JSON
 - Arquivos texto
 - Raspagem de dados XML e HTML
 - Integração com APIs
 - SGBDS
- Limpeza e preparação de dados
 - Lidando com dados ausentes
 - Transformação de dados
 - Manipulação de Strings e expressões regulares
- matplotlib e Seaborn
 - Plotar a partir do shell ou do notebook
 - Interface “MATLAB” e interface orientada a objetos
 - Cores e estilos
 - Rótulos e Título
 - Tipos de gráficos
 - Pontos
 - Barras
 - Error bar
 - Linha
 - Histograma
 - Boxplot
 - Subplots
- Introdução à Machine Learning
 - Tipos de aprendizado
 - Modelos
 - Uso do scikit-learn
 - Classificação