

**Área de Concentração:** 17136

**Criação:** 08/03/2022

**Ativação:** 08/03/2022

**Nr. de Créditos:** 4

**Carga Horária:**

Teórica (por semana)	Prática (por semana)	Estudos (por semana)	Duração	Total
2	2	2	10 semanas	60 horas

**Docente Responsável:**

Ricardo Roberto da Silva

**Objetivos:**

Ilustrar os conceitos básicos de programação e algoritmos para análise exploratória de dados, utilizando ferramentas que possibilitem aos alunos continuar o aprendizado após o curso.

**Justificativa:**

A crescente disponibilidade de dados provenientes de novas tecnologias analíticas tem aumentado a demanda de conhecimentos no armazenamento, processamento e interpretação da informação gerada por essas abordagens. A comunidade científica tem respondido a essa demanda com o desenvolvimento e difusão da estatística computacional para o gerenciamento da informação. O curso Introdução à Ciência dos Dados utiliza Python, uma linguagem de programação muito popular nas ciências dos dados, com uma sintaxe simples. A plataforma Jupyter, para escrita e execução de código será adotada pelo curso, por ser uma plataforma de código aberto que permite a criação e compartilhamento de documentos que contêm código, equações, visualizações e textos narrativos. Entre os exemplos de aplicações estão: preparação dos dados para análise estatística, simulações numéricas, modelagem estatística, visualização de dados, aprendizado de máquinas e muito mais! A proposta do curso é introduzir o uso do Jupyter, apresentando conceitos básicos de programação e ciências dos dados. Esperamos que a apresentação de exemplos práticos possa estimular a incorporação das ferramentas à rotina de pesquisa dos participantes.

**Conteúdo:**

O curso de Introdução à Ciência dos Dados visa familiarizar o aluno com os principais fundamentos e as técnicas utilizadas nas ciências dos dados, com uma abordagem que permita aos alunos generalizar os conceitos de manipulação de dados, aplicação e validação de modelos, representação gráfica e interpretação dos resultados, para que os mesmos possam ser aplicados a variadas demandas analíticas.

Programa Resumido:

- 1 - Introdução à plataforma jupyter
- 2 - Introdução à linguagem de programação python
- 2.1 - Conceitos básicos de programação
- 2.1.1 - Estruturas de dados
- 2.1.2 - Estruturas de controle
- 2.1.3 - Operações lógicas
- 2.1.4 - Funções
- 2.2 - Formatando dados com a biblioteca pandas
- 2.3 - Realizando análises numéricas com a biblioteca numpy
- 2.4 - Introdução à biblioteca sklearn
- 3 - Análise de regressão
- 4 - Análise de classificação
- 5 - Análise de agrupamento
- 6 - Aprendizado estruturado profundo
- 7 - Projeto de conclusão

**Forma de Avaliação:**

Método:

Aulas expositivas, vídeo aulas, aplicação do conceito de sala invertida, onde os alunos assistem às aulas em vídeo e temos uma seção presencial para discutir o que foi estudado e fazer exercícios. A plataforma online (<http://seriema.fcfrp.usp.br/elearn>) conta ainda com exercícios para cada seção da parte prática. No final teremos um projeto de análise de dados, que pode ser proposto pelo estudante analisando dados da sua própria pesquisa.

Critério:

A nota será dada por 75% de exercícios propostos de programação e análise de dados e 25% para o projeto.

**Bibliografia:**

Chollet, F. (2021). Deep learning with Python. Simon and Schuster.  
Eric, M. (2019). Python Crash Course, 2nd Edition. Python Crash Course, 2nd Edition, 544. <https://nostarch.com/pythoncrashcourse2e>

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis: Fifth edition. Cluster Analysis: Fifth Edition, 1–330. <https://doi.org/10.1002/9780470977811>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. <https://doi.org/10.1007/978-1-0716-1418-1>

Mucherino, A., Papajorgji, P., & Pardalos, P. M. (2009). Data mining in agriculture (Vol. 34). Springer Science & Business Media.  
Rogers, Simon, and Mark Girolami. A First Course in Machine Learning. Second Edition. Chapman and Hall/CRC, 2017.

VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc."  
Varmuza, K., & Filzmoser, P. (2016).

Introduction to multivariate statistical analysis in chemometrics. CRC press.

**Idiomas ministrados:**

Português

**Tipo de oferecimento da disciplina:**

Presencial