

Ministério da Educação
Universidade Federal do Paraná
Setor de Tecnologia
Graduação em Engenharia de Produção - Curitiba

**PLANO DE ENSINO
FICHA Nº 2**

Alexandre Checoli Choueiri

Disciplina: Introdução ao Python para Mineração de dados

Código: TP025

Natureza: Optativa

Semestral

Observação: Sugerido ter cursado informática

Co-requisito: não possui

Presencial

C.H. Semestral Total: 60 horas

C.H. Semanal 04

EMENTA

Introdução à linguagem de programação Python. Estruturas de dados, funções. Pacotes Numpy, Pandas, Matplotlib, scikitlearn. Tarefas da mineração de dados. Árvores de decisão. Classificação, Agrupamento. Criação e distribuição (deploy) de programas com interface gráfica usando o pacote Streamlit. Uso de APIs com Python.

PROGRAMA

- Introdução ao Python.
- Variáveis e estruturas de dados.
- Laços de repetição.
- Estruturas condicionais.
- Fatiamento, compreensão de listas.
- Pacote Numpy (arrays, indexação booleana, operações entre arrays).
- Pacote Pandas (Dataframe, leitura e escrita de dados, operações em Dataframes).
- Criação de gráficos com matplotlib (linhas, barras, pizza, histograma, boxplot).
- Tarefas da mineração de dados.
- Classificação/Regressão.
- Árvores de decisão.
- Pacote scikitlearn para a tarefa de classificação.
- Agrupamento.
- Pacote scikitlearn para a tarefa de agrupamento (k-medias).
- Interfaces gráficas com pacote streamlit.
- Uso de APIs com Python.

OBJETIVO GERAL

Capacitar o aluno a utilizar a linguagem de programação Python e suas principais bibliotecas para processamento de dados, bem como para geração dos principais modelos que resolvem tarefas de mineração de dados.

PROCEDIMENTOS DIDÁTICOS

As aulas são divididas em duas partes: parte teórica e parte prática. Na parte teórica, aulas expositivas sobre os conceitos e algoritmos. As aulas serão ministradas presencialmente, com material exposto no quadro e em apresentações de

slides. Nessas aulas serão passados trabalhos em grupo referente ao conteúdo, que farão parte da nota final.

As aulas práticas serão realizadas com computadores, de modo a exercitar os conteúdos teóricos por meio de análises de dados. A linguagem de programação utilizada é o Python, sendo que o ambiente de desenvolvimento fica a critério do aluno (VSCode, PyCharm, Google Colab, Jupyter notebook). Uma breve introdução aos ambientes será realizada, porém as aulas serão focadas nos Notebook Jupyter. Também será dada uma breve introdução ao modo de criar aplicações de mineração de dados na Web usando pacotes próprios do Python, como o streamlit, dessa forma o aluno consegue “produzir” e disponibilizar as suas soluções diretamente no mercado.

FORMAS DE AVALIAÇÃO

A avaliação na disciplina consistirá na realização de a-) atividades semanais, b-) avaliações teóricas individuais, bem como um c-) trabalho final em grupo. As avaliações são descritas em detalhes abaixo:

- a- **Atividades semanais (2 pts):** Atividades referentes aos conteúdos teóricos e práticos, à medida que estes vão sendo passados. As atividades podem começar em sala, porém sempre existe um tempo para entrega (pelo menos 2 dias)
- b- **Avaliações teóricas individuais (3 pts):** Avaliação teórica sobre os conteúdos da disciplina.
- c- **Listas de exercícios (3 pts):** Resolução das listas de exercícios em Python.
- d- **Trabalho final em grupo (4 pts)** Os alunos deverão criar alguma aplicação (qualquer que seja), usando pelo menos um pacote que não foi passado em aula. Explicar o pacote em sala e mostrar o código.

Está prevista a realização de Exame Final para o aluno que obtiver "Média Semestral" igual ou superior a 4,0 e inferior a 7,0, bem como frequência igual ou superior a 75% nas atividades presenciais.

BIBLIOGRAFIA BÁSICA

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. Introduction to data mining. Pearson Education India, 2016.

RUSSELL, Stuart; NORVIG, Peter. Artificial intelligence: a modern approach

2002. CORMEN, Thomas H. et al. Algoritmos: teoria e prática. Editora Campus, v. 2, p. 296, 2002

BIBLIOGRAFIA COMPLEMENTAR

KUBAT, Miroslav. Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7

MITCHEL, Thomas. Machine Learning. McGraw-Hill Science, 1997

ALBON, Chris. Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.", 2018.

AL SWEIGART, Automate the Boring Stuff with Python - Al Sweigart
<https://automatetheboringstuff.com/>

Hal Daumé III, A Course in Machine Learning. <http://ciml.info/>

MATERIAL COMPLEMENTAR

Curso de Machine Learning com ênfase em implementação usando Javascript :
https://www.youtube.com/playlist?list=PLRqwX-V7Uu6YPSwT06y_AEYTqIwbeam3y

COMUNICAÇÃO PROFESSOR ↔ ALUNO

Do professor para os alunos:

- Avisos gerais sobre a disciplina, em particular os referentes à divulgação de materiais didáticos e de atividades, serão informados através de email encaminhado pelo SIGA para o endereço eletrônico cadastrado no SIGA, bem como na página da disciplina: <https://alexandrechecoli.github.io/2-mindados/mindados.html> e durante às aulas.

Dos alunos para o professor:

- Em sala de aula ou pelo email alexandrechecoli@ufpr.br

Professor: Alexandre Checoli Choueiri

Chefe do Departamento: Adriana de Paula Lacerda