



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
SETOR PALOTINA

Departamento de Engenharias e Exatas

Ficha 2 (variável)

Disciplina: Introdução a Ciência de Dados (EaD) Código: DEE645

Natureza:
() Obrigatória (X) Semestral () Anual () Modular
(x) Optativa

Pré-requisito: Co-requisito: Modalidade: () Presencial (x) Totalmente EaD () __ % EaD*

CH Total: 30h						
CH semanal: 2h						
Prática como Componente Curricular (PCC):	Padrão (PD): 30h	Laboratório (LB): 0	Campo (CP): 0	Estágio (ES): 0	Orientada (OR): 0	Prática Específica (PE): 0
Atividade Curricular de Extensão (ACE):						

EMENTA (Unidade Didática)

Introdução à ciência de dados. Descoberta de conhecimento em bases de dados. Exploração de dados: caracterização de dados (objetos e atributos; tipos de dados); dados univariados; dados multivariados; visualização. Pré-processamento: qualidade de dados, limpeza dos dados, valores inconsistentes, objetos redundantes, ruído, outliers, dados desbalanceados. Estudo e projeto de algoritmo preditivo simples: modelagem, desempenho, métricas de avaliação.

JUSTIFICATIVA PARA OFERTA PARCIALMENTE A DISTÂNCIA

PROGRAMA (itens de cada unidade didática)

- Unidade 1: Setup e Fundamentos de Análise de Dados em Python
- Ambientação (UFPR Virtual, Página do Professor, Colab, Jupyter e IDEs).
- Fundamentos de Python: estruturas de dados, funções, manipulação de arquivos e tratamento de exceções.
- As "Big Three" da Ciência de Dados: introdução ao NumPy, Pandas e Matplotlib.
- Unidade 2: Data Wrangling e Análise Exploratória (EDA)
- Operações matemáticas vetorizadas e manipulação de matrizes numéricas.
- Qualidade e Limpeza de Dados: diagnóstico e tratamento de valores ausentes (NaN), imputação, interpolação e tratamento de duplicatas.
- Agregações e Respostas: agrupamentos (*groupby*), filtragens avançadas e tabelas dinâmicas (*pivot tables*).
- A Arte da Visualização de Dados: gráficos estatísticos, mapas de calor (*heatmaps*), correlações, visualização geoespacial interativa (Plotly) e ética visual.
- Unidade 3: Fundamentos de Machine Learning e Regressão
- Tipos de aprendizado (Supervisionado e Não Supervisionado) e o custo do erro.
- *Feature Selection*, *Data Augmentation* e o risco da Maldição da Dimensionalidade.
- Validação de Modelos: *Train/Test Split*, *K-Fold Cross-Validation*, Matriz de Confusão e Métricas de Desempenho.
- Modelos de Regressão: baseline, Regressão Linear Múltipla, análise de resíduos, *Pipelines* e Regularização (Ridge e Lasso).
- Unidade 4: Modelagem Preditiva e Classificação
- Algoritmos Clássicos baseados em distância e probabilidade: KNN, Naive Bayes e Máquinas de Vetores de Suporte (SVM).
- Classificadores interpretáveis: Regressão Logística e Árvores de Decisão.
- *Ensembles* de Alta Performance: Random Forest e XGBoost.
- Aprendizado Não Supervisionado: Clusterização e zonas de manejo com K-Means e redução de dimensionalidade (PCA).
- Unidade 5: Tópicos Avançados e Produto Final (Deploy)
- Introdução a Redes Neurais Profundas (Deep Learning).
- Visão Computacional aplicada: detecção e segmentação de imagens no contexto agrícola (YOLO, U-Net).
- Transformando Modelos em Aplicações: criação de interfaces web com Gradio e Streamlit.
- Desenvolvimento de projeto prático integrador de ponta a ponta.

OBJETIVO GERAL

Capacitar o aluno a extrair, limpar, analisar e modelar dados utilizando a linguagem Python, desenvolvendo soluções completas de Machine Learning, desde a análise exploratória até o deploy de aplicações web interativas, com foco na resolução de problemas reais.

OBJETIVO ESPECÍFICO

- Apresentar o ecossistema moderno de Ciência de Dados em Python (NumPy, Pandas, Scikit-Learn, Plotly).
- Ensinar técnicas robustas de *Data Wrangling* para lidar com dados ruidosos, faltantes e inconsistentes do mundo real.
- Desenvolver a capacidade de criar visualizações de dados claras, interativas e eticamente corretas para tomada de decisão.
- Projetar, treinar e validar com rigor algoritmos preditivos de regressão e classificação (modelos clássicos e *ensembles*).
- Introduzir conceitos modernos de IA, como Visão Computacional e Deep Learning, aplicados a cenários práticos (ex: Agro 5.0).
- Capacitar o estudante a empacotar seus modelos preditivos em interfaces web amigáveis para usuários finais.

PROCEDIMENTOS DIDÁTICOS

Metodologia de Ensino: A disciplina é desenvolvida na modalidade EaD (assíncrona), onde cada aluno pode seguir o seu ritmo (*self-pacing* - ritmo próprio). Considerando que a carga horária de 30h é enxuta para o volume e a densidade do conteúdo de Ciência de Dados, a metodologia será fortemente baseada em **Notebooks interativos no Google Colab**. Este formato garante que o aluno consiga absorver o material de forma prática e aplicada, executando códigos na nuvem sem ficar travado com instalações complexas de ferramentas ou lidando com teorias excessivamente longas.

Ambiente Virtual e Materiais Didáticos: O conteúdo completo da disciplina (textos, notebooks, vídeos e links) está hospedado e organizado na página oficial do professor. No entanto, **toda a interação, avisos e centralização do curso ocorrerão obrigatoriamente no ambiente da UFPR Virtual**. Diferentes recursos didáticos serão utilizados, incluindo gravações em vídeo, artigos, links para leitura e formulários práticos.

Sistema de Comunicação e Tutoria: A comunicação oficial e a interação sobre a disciplina ocorrerão exclusivamente pelo ambiente da UFPR Virtual (fóruns e murais de avisos). O e-mail institucional será reservado apenas para informações administrativas, recados e lembretes. A tutoria para esclarecimento de dúvidas será realizada de forma assíncrona pelos fóruns ou de forma síncrona/individualizada mediante agendamento prévio (via Microsoft Teams, que é a ferramenta Institucional).

Dinâmica de Prazos e Controle de Frequência: Embora a disciplina seja EaD e assíncrona, ela é planejada para acontecer rigorosamente dentro do intervalo do semestre letivo, com início e fim bem definidos. Para garantir o bom andamento da turma, evitar que o aluno deixe as atividades para a última hora (o que inviabiliza a aprendizagem e a participação) e permitir a melhor gestão da disciplina pelo professor, foi estabelecida a seguinte regra de progressão: **após o primeiro mês de início do semestre, a cada nova semana será fechado o acesso a um dos módulos iniciais sequencialmente**. Sendo que a totalidade dos módulos será encerrada no último dia letivo do semestre.

Ao final de cada aula/módulo, o aluno deverá responder a um **formulário de perguntas**. Este formulário tem dupla função:

1. Praticar os conceitos, auxiliar na fixação do conteúdo e avaliar a aprendizagem.
2. Contabilizar a **presença e participação** do aluno naquele módulo específico.

Infraestrutura Tecnológica e Acessibilidade: Os alunos poderão verificar o agendamento para o uso de laboratório de informática do Setor Palotina (disponibilizado pelo Técnico em TI do Setor) para executar atividades das aulas ou estudar. No aspecto de acessibilidade, os vídeos fornecidos oferecerão a opção de configurar legendas automáticas. Em caso de outras necessidades especiais, o aluno deve consultar o professor.

FORMAS DE AVALIAÇÃO

A avaliação será composta por:

Avaliação das atividades semanais (AS) realizadas no decorrer da disciplina (formulários);
Avaliação do projeto final (PF) da disciplina.

Dessa forma, a nota final (NF) será obtida através de:

$$NF = (AS*0.5)+(PF*0.5)$$

BIBLIOGRAFIA BÁSICA

COPPIN, B. Inteligência Artificial. Rio de Janeiro: LTC, 2010.

MENEZES, N. N. C. Introdução à Programação com Python. Novatec, 2010.

BARRY, P.; GRIFFITHS D. Use a Cabeça! Programação. Alta Books, 2010.

BIBLIOGRAFIA COMPLEMENTAR

MORETIN, L. G. Estatística básica. São Paulo: Pearson Education, 2009.

DOWNEY, A.; ELKNER, J.; MEYERS, C. Como Pensar como um Cientista da Computação usando Python. 2009. Disponível

em: <http://www3.ifrn.edu.br/~jurandy/fdp/doc/aprendapython/index.html>

RUSSELL, S., NORVIG, P. Artificial Intelligence: A Modern Approach. 4a ed .Editora Pearson. 2020. p. 1132.

CHOLLET, F. Deep Learning with Python. New York: Manning, 2018.

DATA SCIENCE ACADEMY. Deep Learning Book: em Português, Online e Grátis. 2019.

Disponível em: <<http://deeplearningbook.com.br/>>.