* **Título da aba do navegador:** INF 493 - Ciência de Dados
* **Texto do cabeçalho:** Informações Gerais de Disciplina
* **Título da página:** INF 493 - Tópicos Especiais III: Ciência de Dados
* **Informações Gerais:**
  + **Centro:** Ciências exatas (url: <https://cce.ufv.br/>)
  + **Departamento:** Departamento de Informática (url: <https://www2.dpi.ufv.br/>)
  + **Currículo:** 2020 (url: <https://ccp.ufv.br/matriz-2020-ou-posterior/>)
  + **Créditos:** 4
  + **Carga horária total:** 60h
  + **Carga horária teórica semanal:** 4h
  + **Carga horária prática semanal:** 0h
  + **Semestre de oferta:** II
  + **Obrigatória:** Não
* **Objetivos:**
  + Prover ao estudante uma visão geral da área de Ciência de Dados (Data Science), com foco nas etapas de preparação de dados, análise de resultados e modelos e em aplicações reais ou realísticas. Relacionar a visão ampla do conceito de Ciência de Dados com as áreas correlatas de Aprendizado de Máquina e Mineração de Dados. Fornecer ao estudante um repertório de técnicas que o possibilite atuar em contextos envolvendo suporte à tomada de decisões, seja na área acadêmica ou não.
* **Ementa:**
  + Introdução e terminologia básica. Fundamentos de análise de dados. Fundamentos de visualização de dados. Técnicas de pré-processamento e preparação de dados. Redução de dimensionalidade e seleção de atributos. Análise de Agrupamentos. Classificação. Regressão. Comitês. Explicabilidade de modelos preditivos. Outras técnicas relacionadas. Áreas de aplicação modernas.
* **Pré-requisitos:**
  + INF 213 - Estruturas de Dados
  + EST 105 - Iniciação à Estatística
* **Co-requisitos:**
  + Não há
* **Conteúdo Programático:**
  + **Introdução**
    - O que é Ciência de Dados (Data Science)
    - Conceitos importantes: Data Mining versus Machine Learning versus Data Science versus Inteligência Artificial
    - Visão geral do curso e suas lições
  + **Fundamentos de análise de dados**
    - Representação de Dados
    - Tipos de atributos
    - Fundamentos de Estatística
    - Visualização de dados
    - Análise Uni e Multivariada
    - Introdução aos testes de hipótese
  + **Pré-processamento e preparação de dados**
    - Normalização (escalonamento e padronização)
    - Identificação e remoção de dados discrepantes
    - Tratamento de valores faltantes
    - Categorização de dados numéricos
    - Pré-processamento para dados categóricos
  + **Principais tarefas de mineração de dados e aprendizado de máquina**
    - Agrupamentos (clusters)
    - Classificação
    - Regressão
    - Metodologias de treinamentos
    - Principais algoritmos
    - Métricas de avaliação
    - Comparação de modelos
    - Validação Cruzada
    - ANOVA
    - Casos especiais
      * Overfitting e Underfitting
      * Classes raras
      * Multirrótulo
      * Outros
  + **Seleção de atributos e redução de dimensionalidade**
  + **Comitês**
  + **Explicabilidade de modelos**
  + **Outros tipos de aprendizado**
    - Semi-supervisionado
    - Séries temporais
    - Aprendizado por reforço
    - Regras de Associação
  + **Algumas áreas de aplicação modernas**
    - Processamento de linguagem natural
    - Visão computacional
    - Modelos generativos
    - Federated Learning
* **Avaliações**

| **Avaliações** | | |
| --- | --- | --- |
| **Tipo** | **Percentual da Nota** | **Quantidade** |
| **Questionário** | 6% | 5 |
| **Trabalho prático** | 10% | 3 |
| **Projeto Final** | 40% | 1 |