
CURSO: Ciência de Dados e Inteligência Artificial – 1º semestre de 2024
DISCIPLINA: Projetos especiais
PROFESSOR(ES): Thiago Guerrera Martins
CARGA HORÁRIA: 60h
PRÉ-REQUISITO: conhecimentos básicos de SQL, e familiaridade com conceitos de engenharia de dados e análise de dados.
HORÁRIO E SALA DE ATENDIMENTO: Terça-feira de 08:30 às 09:30
SALA: Zoom

PLANO DE ENSINO

1. Ementa

O curso é dividido em dois módulos, centrado ao redor de um problema prático escolhido pelo professor:

Módulo 1: Organização de Dados com dbt (30 horas)

1. Introdução à Engenharia de Dados com dbt
 - Conceitos de engenharia de dados
 - O papel do dbt na orquestração de pipelines de dados
 - Configuração do ambiente de desenvolvimento dbt
2. Modelagem de Dados com dbt
 - Definição de modelos em dbt
 - Transformações de dados com SQL
 - Versionamento de modelos
3. Organização de Dados no dbt Cloud
 - Uso do dbt Cloud para gerenciamento de projetos
 - Agendamento e automação de jobs de dbt
 - Integração com outras ferramentas de análise
4. Manipulação e Limpeza de Dados

-
- Transformação de dados com SQL no dbt
 - Resolução de problemas comuns de qualidade de dados
 - Trabalhar com diferentes tipos de fontes de dados
5. Colaboração e Controle de Versão em Projetos dbt
- Estratégias de colaboração em projetos dbt
 - Controle de versão de modelos e transformações
 - Boas práticas de desenvolvimento

Módulo 2: Análise e Modelagem de Dados Avançada (30 horas)

Nesse módulo os alunos deverão utilizar os dados organizados no módulo 1 e criar modelos para resolver o problema específico proposto pelo professor. Entre os modelos que os alunos serão expostos estão:

- Generalized Additive Models
- Deep Learning
- Tree-based models

Entre as bibliotecas que os alunos deverão se expor estão:

- Scikit-learning
- TensorFlow ou PyTorch
- XGBoost

2. Objetivos da disciplina

No módulo 1 os alunos aprenderão a organizar, transformar e gerenciar dados utilizando a ferramenta dbt (Data Build Tool), que desempenha um papel fundamental na engenharia de dados e análise de dados.

O módulo 2 tem como objetivo aprofundar o conhecimento dos alunos em análise de dados e modelagem estatística, utilizando as bibliotecas e modelos mais comuns e avançados.

3. Procedimentos de ensino (metodologia)

Sessões síncronas presenciais.

4. Conteúdo programático detalhado

| Datas | Tópico | Atividades |
|-----------|---|------------|
| Semana 1 | Introdução à Engenharia de Dados com dbt | |
| Semana 2 | Conceitos de engenharia de dados | |
| Semana 3 | Modelagem de Dados com dbt | |
| Semana 4 | Configuração do ambiente de desenvolvimento dbt | |
| Semana 5 | Transformações de dados com SQL no dbt | |
| Semana 6 | Versionamento de modelos | |
| Semana 7 | Organização de Dados no dbt Cloud | |
| Semana 8 | Manipulação e Limpeza de Dados | |
| Semana 9 | Generalized Additive Models (GAM) - Parte 1 | |
| Semana 10 | Generalized Additive Models (GAM) - Parte 2 | |
| Semana 11 | Generalized Additive Models (GAM) - Parte 3 | |
| Semana 12 | Deep Learning - Parte 1 | |
| Semana 13 | Deep Learning - Parte 2 | |
| Semana 14 | Deep Learning - Parte 3 | |
| Semana 15 | Tree-based Models - Parte 1 | |
| Semana 16 | Tree-based Models - Parte 2 | |
| Semana 17 | Tree-based Models - Parte 3 | |

5. Procedimentos de avaliação

Os alunos serão avaliados através de dois trabalhos, a compor duas notas P1 e P2. Caso o aluno não consiga aprovação, poderá realizar uma nova avaliação que substituirá obrigatoriamente a menor entre P1 e P2.

6. Bibliografia Obrigatória

- Notas de aula e notas computacionais elaboradas pelo professor e disponibilizadas ao longo das aulas.