

Data Mining e Graph Mining

Inteligência Artificial

Professor André Luiz Esperidião

APRESENTAÇÕES

BREVE APRESENTAÇÃO

Professor André Luiz Esperidião

- ☐ Graduação em Tecnologia em Processamento de Dados
- ☐ Especialização em Análise e Projeto de Sistemas
- ☐ MBA em Gestão de Projetos com PMI-PMBOK
- ☐ MBA em Gestão Empresarial
- ☐ Mestrado em Engenharia Elétrica e de Computação
Linha de pesquisa: Sistemas Inteligentes
- ☐ Doutorado em Engenharia Elétrica
Linha de pesquisa: Machine Learning

✓ Certificações: ITIL®, COBIT®, ISO 27002, Green Belt 6 Sigma.

BREVE APRESENTAÇÃO

Apresentação dos alunos (individual)

- ☐ Qual o seu nome?
- ☐ Razão da escolha do curso?
- ☐ Atua ou já atuou na área de tecnologia?
- ☐ O que você sabe sobre Data Mining e Graph Mining?
- ☐ Qual a sua expectativa sobre essa disciplina?

BREVE APRESENTAÇÃO

Apresentação dos alunos (coletivo)

☐ Já tiveram contato com Estatística?

☐ Já tiveram contato com Algoritmos e Estruturas de Dados?

A DISCIPLINA

Data Mining e Graph Mining

A DISCIPLINA DATA MINING E GRAPH MINING

Módulo: COMPUTAÇÃO DISTRIBUÍDA

Curso: Tecnologia em Inteligência Artificial

Unidade Curricular: Data Mining e Graph Mining

Carga Horária: 120h

Objetivo Geral: Desenvolver fundamentos técnicos e científicos referente à mineração de dados e mineração de dados em estruturas de dados do tipo grafo, bem como capacidades sociais, organizativas e metodológicas, de acordo com a área de ocupação no mundo do trabalho.

A DISCIPLINA DATA MINING E GRAPH MINING

“Hoje quase tudo vira dado.
Pouca coisa vira decisão.”

Por que essa disciplina existe?

- Onde Data Mining aparece no mundo real?
- Por que grafos são diferentes?
 - 💣 “Tabela não explica rede. E rede explica muito mais do que parece.”

A DISCIPLINA DATA MINING E GRAPH MINING



ORGANIZAÇÃO SEMANAL DA DISCIPLINA (120h)



Estrutura fixa semanal

- **Quarta – feira:** 2h (20:45h às 22:45h)
- **Sexta – feira:** 4h (18:30h às 20:30h – 20:45h às 22:45h)



Total: 6h/semana



Duração do curso: $120h \div 6h = 20$ semanas

A DISCIPLINA DATA MINING E GRAPH MINING

Etapas	Tema	C/H
1	Fundamentos de Data Mining	25h
2	Preparação de Dados	20h
3	Modelagem e Avaliação de Modelos	25h
4	Graph Mining na Prática	25h
5	Graph Mining em Ambiente Real	15h
6	Projeto Integrador	10h
Total		120h

A DISCIPLINA DATA MINING E GRAPH MINING

1 – FUNDAMENTOS OPERACIONAIS DE DATA MINING (20h)

- Entender problema → transformar em tarefa de DM
- CRISP-DM aplicado (checklist operacional)
- Leitura e inspeção de datasets reais
- Primeira modelagem (classificação simples)



Lab:

- Dataset real

A DISCIPLINA DATA MINING E GRAPH MINING

2 – PREPARAÇÃO DE DADOS E FEATURE ENGINEERING (25h)

- Limpeza de dados reais
- Encoding de categorias
- Normalização
- Seleção de características



Lab:

- Dataset real sabotado
- Comparação antes e depois

A DISCIPLINA DATA MINING E GRAPH MINING

3 – MODELAGEM E AVALIAÇÃO DE MODELOS (25h)

- Classificação vs regressão vs clusterização
- Comparar modelos
- Ajuste de hiperparâmetros
- Métricas corretas para cada problema



Lab:

- Mesmo dataset, modelos diferentes
- Escolha justificada

A DISCIPLINA DATA MINING E GRAPH MINING

4 – GRAPH MINING NA PRÁTICA (25h)

- Modelar dados como grafo
- Construir grafo a partir de dados tabulares
- Calcular métricas estruturais
- Detectar comunidades
- Analisar influência



Lab:

- Simulação
- Visualização e interpretação

A DISCIPLINA DATA MINING E GRAPH MINING

5 – GRAPH MINING EM AMBIENTE REAL (15h)

- Grafos grandes (limitações reais)
- Introdução a Neo4j
- Introdução a Spark GraphFrames
- Trade-offs de escala



Lab:

- Subgrafo grande
- Gargalos de processamento
- Discussão de arquitetura

A DISCIPLINA DATA MINING E GRAPH MINING

6 – Projeto integrador (10h)

- Problema real
- Pipeline completo
- Escolha a técnica
- Resultados



Lab:

- Técnica
- Organização
- Comunicação
- Tomada de Decisão

STACK ESSENCIAL DE FERRAMENTAS

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 1 – BASE OBRIGATÓRIA (curso prático)

1. Python

 Motivo: padrão de mercado + ecossistema maduro.

- Linguagem principal do curso
- Foco em leitura, transformação e modelagem

Bibliotecas:

- pandas → manipulação de dados
- numpy → base numérica
- scikit-learn → modelos clássicos
- matplotlib / seaborn → visualização básica

 Observação didática:

- Aprenda **pipeline**.

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 1 – BASE OBRIGATÓRIA (curso prático - continuação)

2. Jupyter Notebook

 Motivo: aprendizado incremental + visual.

- Excelente para:
 - exploração
 - experimentos
 - relatórios técnicos

 Atenção:

- Notebook não é produção.

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 2 – GRAPH MINING (analítico)

3. NetworkX



Motivo: clareza conceitual.

- Construção de grafos
- Métricas estruturais
- Algoritmos clássicos



Atenção:

- NetworkX é perfeito para aprender... e péssimo para produção.
E tudo bem..

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 2 – GRAPH MINING (analítico)

4. Ferramenta de Visualização de Grafos

Opções:

- matplotlib (simples)
- pyvis (interativo)
- Gephi (visual, fora do Python)

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 3 – GRAPH MINING EM AMBIENTE REAL

5. Neo4j (Graph Database)



Motivo: banco de grafos mais usado no mercado.

- Modelo Property Graph
- Linguagem Cypher
- Ideal para:
 - consultas complexas
 - grafos dinâmicos



Aprender: conceito, modelagem, consultas essenciais

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 3 – GRAPH MINING EM AMBIENTE REAL (continuação)

6. Apache Spark + GraphFrames

 Motivo: escala.

- Processamento distribuído
- Grafos grandes
- Integração com Data Mining

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 4 – INFRAESTRUTURA DIDÁTICA

7. Google Colab ou Ambiente Local

- Colab:
 - fácil
 - sem instalação
- Local:
 - Mais controle

STACK ESSENCIAL DE FERRAMENTAS

CAMADA 4 – INFRAESTRUTURA DIDÁTICA (continuação)

8. Git (opcional, mas recomendado)



Motivo: competência profissional.

- Versionamento
- Organização de projeto
- Trabalho em grupo

STACK ESSENCIAL DE FERRAMENTAS

STACK FINAL (RESUMO EXECUTIVO)

Finalidade	Ferramenta
Data Mining	Python + pandas + sklearn
Exploração	Jupyter
Graph Mining	NetworkX
Visualização	Gephi ou pyvis
Grafos reais	Neo4j
Escala	Spark + GraphFrames
Ambiente	Colab / Local
Profissional	Git

BIBLIOGRAFIA

BIBLIOGRAFIA



BIBLIOGRAFIA BÁSICA

- AMARAL, Fernando. Aprenda Mineração de Dados. Rio de Janeiro: Editora Alta Books, 2019. E-book. ISBN 9786555206852. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9786555206852/>. Acesso em: 29 jul. 2024.
- CASTRO, Daniel Gomes Ferrari Leandro Nunes de. Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações. São Paulo: SRV Editora LTDA, 2016. E-book. ISBN 978-85-472-0100-5. Disponível em: <https://app.minhabiblioteca.com.br/#/books/978-85-472-0100-5/>. Acesso em: 29 jul. 2024.
- GOLDSCHMIDT, Ronaldo. Data Mining. Rio de Janeiro: Grupo GEN, 2015. E-book. ISBN 9788595156395. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788595156395/>. Acesso em: 29 jul. 2024.

BIBLIOGRAFIA



BIBLIOGRAFIA COMPLEMENTAR

- GOLDBARG, Marco. Grafos. Rio de Janeiro: Grupo GEN, 2012. E-book. ISBN 9788595155756. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788595155756/>. Acesso em: 29 jul. 2024.
- MARIANO, Diego César B.; MARQUES, Leonardo T.; SILVA, Marcel S.; et al. Data Mining. Porto Alegre: Grupo A, 2021. E-book. ISBN 9786556900292. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9786556900292/>. Acesso em: 29 jul. 2024.
- SILVA, Leandro Augusto da; PERES, Sarajane M.; BOSCARIOLI, Clodis. Introdução à Mineração de Dados - Com Aplicações em R. Rio de Janeiro: Grupo GEN, 2016. E-book. ISBN 9788595155473. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788595155473/>. Acesso em: 29 jul. 2024.

BIBLIOGRAFIA



BIBLIOGRAFIA COMPLEMENTAR (CONTINUAÇÃO)

- SZWARCFITER, Jayme L. Teoria Computacional de Grafos - Os Algoritmos. Rio de Janeiro: Grupo GEN, 2018. E-book. ISBN 9788595155183. Disponível em: <https://app.minhabiblioteca.com.br/#/books/9788595155183/>. Acesso em: 29 jul. 2024.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. Introdução ao datamining mineração de dados. Rio de Janeiro: Unama - Universidade da Amazônia, 2009. 900 p. ISBN 978-85-7393-761-9.

Data Mining e Graph Mining

Inteligência Artificial

Professor André Luiz Esperidião