



FICHA DE COMPONENTE CURRICULAR

CÓDIGO: GBC212	COMPONENTE CURRICULAR: MINERAÇÃO DE DADOS	
UNIDADE ACADÊMICA OFERTANTE: FACULDADE DE COMPUTAÇÃO		SIGLA: FACOM
CH TOTAL TEÓRICA: 60 horas	CH TOTAL PRÁTICA: 00 horas	CH TOTAL: 60 horas

1. OBJETIVOS

Discutir sobre as principais tarefas e técnicas de Mineração de Dados.
Aplicar ferramentas de Mineração de Dados em problemas práticos.
Implementar suas próprias ferramentas de Mineração de Dados.

2. EMENTA

Introdução e Motivação ao Processo de Descoberta de Conhecimento em Bases de Dados (KDD). Etapas do Processo de KDD. Técnicas de Pré-processamento dos Dados. Tarefas, Algoritmos e Paradigmas de Mineração de Dados: Associações, Classificação, Agrupamentos, Detecção de Outliers. Pós-processamento dos Resultados: Análise, Interpretação e Visualização. Ferramentas de Mineração de Dados.

3. PROGRAMA

1. Introdução

- O que é Mineração de Dados - o que é Descoberta de Conhecimento (KDD)
- As fases do processo de KDD : pré-processamento dos dados (limpeza, transformação, seleção de atributos), mineração dos dados, pós-processamento dos resultados (análise, interpretação, visualização)
- Principais Tarefas de Mineração de Dados.

2. Preparação dos Dados

- Sumarização dos dados: medindo a tendência central e a dispersão dos dados
- Limpeza dos dados: valores ausentes, tratamento de ruídos
- Integração e Transformação dos dados
- Redução dos dados: seleção de atributos, redução de dimensionalidade.
- Discretização, Normalização.

3. Associação

- Mineração de Regras de Associação – Algoritmo Apriori e variantes
- Mineração de Sequências – Algoritmos GSP e Prefix-Span.

4. **Classificação**

- O que é um classificador
- Arvore de Decisão
- Classificadores baseados nos vizinhos mais próximos (KNN)
- Classificadores baseados em Redes Bayesianas de Crença
- Classificadores baseados em Redes Neurais
- Avaliação de Performance: Método Holdout, Cross-Validation, Bootstrap
- Técnicas de classificação para classes não-balanceadas: métricas alternativas, precisão, recall, Curva Roc.

5. **Agrupamentos (*clusters*)**

- Diferentes tipos de *clusters* : bem separados, baseados em centróides, baseados em grafos, baseados em densidade.
- Diferentes tipos de *clusterização* : por particionamento, hierárquico, exclusivos, fuzzy, com justaposição, completas, parciais
- Método K-Means e K-Medóides – análise de performance, complexidade
- Método hierárquico aglomerativo – análise de performance, complexidade
- Método DBSCAN – análise de performance, complexidade

6. **Análise de *Clusters***

- medidas: coesão, separação, SSE, coeficiente de silhueta
- técnicas para determinar o número correto de clusters
- técnicas para determinar a tendência de clusters nos dados

7. **Detecção de Anomalias (*Outliers*)**

- Introdução: causas de anomalias
- Técnicas para detecção de anomalias: estatísticas e baseadas em proximidade

8. **Pós-Processamento: Análise, Interpretação e Visualização.**

9. **Aplicações: utilização de ferramentas de Mineração de Dados (ao longo de todo o componente curricular)**

4. **BIBLIOGRAFIA BÁSICA**

1. TAN, Pang-Ning. **Introdução ao Datamining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.
2. AMARAL, Fernando. **Aprenda mineração de dados**: teoria e prática_. Rio de Janeiro: Alta Books, 2019.
3. CASTRO, Leandro Nunes de. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

5. **BIBLIOGRAFIA COMPLEMENTAR**

1. CARVALHO, Luís Alfredo Vidal de. **Datamining**: a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Ciência Moderna, 2005.
2. SILVA, Leandro Augusto da. **Introdução à mineração de dados**: com aplicações em R. Rio de Janeiro: GEN LTC, 2016.
3. HAN, Jiawei. **Data mining**: concepts and techniques. 2nd ed. Amsterdam; Boston; San Francisco: Morgan Kaufmann, c2006.
4. WITTEN, I. H. **Data mining**: practical machine learning tools and techniques with Java implementations. San Francisco: Morgan Kaufman, c2000.
5. KANTARDZIC, Mehmed. **Data mining**: concepts, models, methods, and algorithms. 2. ed. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, 2011.

6. APROVAÇÃO

Maria Adriana Vidigal de Lima
Coordenadora do Curso de Ciência da
Computação

Maurício Cunha Escarpinati
Diretor da Faculdade de
Computação



Documento assinado eletronicamente por **Maria Adriana Vidigal de Lima, Coordenador(a)**, em 26/01/2024, às 15:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Diretor(a)**, em 19/02/2024, às 11:33, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5129248** e o código CRC **56237664**.