



UNIVERSIDADE FEDERAL DE UBERLÂNDIA



FICHA DE COMPONENTE CURRICULAR

CÓDIGO: FACOM33504	COMPONENTE CURRICULAR: Ciência de Dados	
UNIDADE ACADÊMICA OFERTANTE: Faculdade de Computação		SIGLA: FACOM
CH TOTAL TEÓRICA: 45 horas	CH TOTAL PRÁTICA: 15 horas	CH TOTAL: 60 horas

1. **OBJETIVOS**

Geral: O objetivo desta disciplina é introduzir o aluno às principais tarefas e técnicas da Ciência de Dados, visando desenvolver a sua habilidade na aplicação dessas ferramentas em problemas práticos e na implementação das suas próprias ferramentas.

Específicos:

- Conceituar dados e conhecimento;
- Discutir as fases da descoberta do conhecimento por meio de dados (KDD);
- Estudar os conceitos, técnicas e aplicações de associação, classificação e agrupamento;
- Capacitar o aluno a efetuar análises, interpretação e visualização 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 à Ciência de Dados
 1. O que é Ciência de Dados?
 2. As fases do processo de Descoberta de Conhecimento (KDD) : pré-processamento dos dados (limpeza, transformação, seleção de atributos), mineração dos dados, pós-processamento dos resultados (análise, interpretação, visualização)
 3. Principais tarefas de Ciência de Dados
2. Preparação dos Dados

1. Sumarização dos dados: medindo a tendência central e a dispersão dos dados
 2. Limpeza dos dados: valores ausentes, tratamento de ruídos
 3. Integração e Transformação dos dados
 4. Redução dos dados: seleção de atributos, redução de dimensionalidade.
 5. Discretização, Normalização.
3. Associação
 1. Regras de Associação — Algoritmo Apriori e variantes
 2. Mineração de Sequências—Algoritmos GSP e Prefix-Span
 4. Classificação de dados
 1. O que é um classificador?
 2. Classificadores baseados nos vizinhos mais próximos (KNN)
 3. Classificadores baseados em Redes Bayesianas de Crença
 4. Avaliação de Performance: Método Holdout, Cross-Validation, Bootstrap
 5. Técnicas de classificação para classes não balanceadas: métricas alternativas, precisão, recall, Curva ROC
 5. Agrupamento de dados
 1. Diferentes tipos de agrupamentos (clusters): bem separados, baseados em centróides, baseados em grafos, baseados em densidade.
 2. Diferentes tipos de clusterização: por particionamento, hierárquico, exclusivos, fuzzy, com justaposição, completas, parciais
 3. Método K-Means — análise de performance, complexidade
 4. Método hierárquico aglomerativo — análise de performance, complexidade
 5. Método DBSCAN — análise de performance, complexidade
 6. Análise de Agrupamentos
 1. Medidas: coesão, separação, SSE, coeficiente de silhueta
 2. Técnicas para determinar o número correto de clusters
 3. Técnicas para determinar a tendência de clusters nos dados
 7. Detecção de Anomalias (Outliers)
 1. Introdução: causas de anomalias
 2. Técnicas para detecção de anomalias: estatísticas e baseadas em proximidade
 8. Pós-Processamento dos resultados
 1. Análise e Interpretação dos resultados
 2. Visualização dos resultados
 9. Tópicos especiais em Ciências de Dados
 1. Ferramentas, repositórios e plataformas disponíveis
 2. Aplicações e estudos de caso

4. BIBLIOGRAFIA BÁSICA

BISHOP, Christopher M. **Pattern recognition and machine learning**. New York: Springer, 2006.

TAN, Pang-Ning, STEINBACH, Michael, KUMAR, Vipin. **Introdução ao datamining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. **Data mining**: practical machine learning tools and techniques. 3rd. ed. Boston: Morgan Kaufman, 2011.

5. BIBLIOGRAFIA COMPLEMENTAR

DUDA, Richard O.; HART, Peter E.; STORK, David G. **Pattern classification**. 2nd. ed. New York: John Wiley & Sons, 2001.

HAN, Jiawei; KAMBER, Micheline. **Data mining**: concepts and techniques. 2nd. ed. Boston: Morgan Kaufmann, 2006.

MITCHELL, Tom M. **Machine learning**. Chennai: McGraw-Hill Education, 1997.

RUSSELL, Matthew A. **Mineração de dados da web social**. São Paulo: Novatec, 2011.

TORGO, Luís. **Data mining with R**: learning with case studies. Boca Raton: Chapman & Hall/CRC Press, 2011.

6. APROVAÇÃO

Prof. Dr. Rafael Dias Araújo
Coordenador do Curso de Sistemas de Informação
Campus Monte Carmelo

Prof. Dr. Mauricio Cunha Escarpinati
Diretor da Faculdade de Computação



Documento assinado eletronicamente por **Rafael Dias Araújo, Coordenador(a)**, em 25/01/2023, às 22:13, 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 26/01/2023, às 17:18, 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 **4043787** e o código CRC **77F199BF**.