



UNIVERSIDADE FEDERAL DE UBERLÂNDIA



FICHA DE COMPONENTE CURRICULAR

CÓDIGO:	COMPONENTE CURRICULAR: MINERAÇÃO DE DADOS	
UNIDADE ACADÊMICA OFERTANTE: FACULDADE DE COMPUTAÇÃO		SIGLA: FACOM
CH TOTAL TEÓRICA: 30 horas	CH TOTAL PRÁTICA: 30 horas	CH TOTAL: 60 horas

1. **OBJETIVOS**

Os objetivos da disciplina são: Introduzir o aluno às principais tarefas e técnicas de Mineração de Dados. Habilitar o aluno a aplicar ferramentas de Mineração de Dados em problemas práticos. Habilitar o aluno a 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 Pre-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**

1.1. O que é Mineração de Dados - o que é Descoberta de Conhecimento (KDD).

1.2. 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).

1.3. Principais Tarefas de Mineração de Dados.

2. Preparação dos Dados

2.1. Sumarização dos dados: medindo a tendência central e a dispersão dos dados.

2.2. Limpeza dos dados: valores ausentes, tratamento de ruídos.

2.3. Integração e Transformação dos dados.

2.4. Redução dos dados: seleção de atributos, redução de dimensionalidade.

2.5. Discretização, Normalização.

3. Associação

3.1. Mineração de Regras de Associação – Algoritmo Apriori e variantes.

3.2. Mineração de Sequências – Algoritmos GSP e Prefix-Span.

4. Classificação

- 4.1. O que é um classificador?
- 4.2. Árvore de Decisão.
- 4.3. Classificadores baseados nos vizinhos mais próximos (KNN).
- 4.4. Classificadores baseados em Redes Bayesianas de Crença.
- 4.5. Avaliação de Performance: Método Holdout, Cross-Validation, Bootstrap.

5. Agrupamentos (Clusters)

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

6. Análise de Clusters

- 6.1. Medidas: coesão, separação, SSE, coeficiente de silhueta.
- 6.2. Técnicas para determinar o número correto de clusters.
- 6.3. Técnicas para determinar a tendência de clusters nos dados.

7. Detecção de Anomalias (Outliers)

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

8. Pós-Processamento: Análise, Interpretação e Visualização**4. BIBLIOGRAFIA BÁSICA**

1. TAN, Pang-Ning. **Introdução ao Data Mining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.
2. J. Han, M. Kamber: **Data Mining: Concepts and Techniques**, 2a. Ed., Morgan Kaufmann, 2006.
3. I. H. Witten, E. Frank: **Data Mining: Practical Machine Learning Tools and Techniques**, Morgan Kaufmann, 2005.

5. BIBLIOGRAFIA COMPLEMENTAR

1. CHAKRABARTI, Soumen. **Mining the Web**: discovering knowledge from hypertext data. San Francisco: Morgan Kaufman, c2003.
2. KANTARDZIC, Mehmed. **Data mining**: concepts, models, methods, and algorithms. 2nd ed. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2011. E-Book.
3. LAROSE, Daniel T. **Data mining methods and models**. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2006. E-Book.
4. JENSEN, Finn V. **Bayesian networks and decision graphs**. 2nd ed. New York: Springer, c2007.
5. CHERKASSKY, Vladimir S. **Learning from data**: concepts, theory, and methods. 2nd ed. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2007. E-Book.

6. APROVAÇÃO

Prof. Dr. Jean Carlos Domingos
Coordenador do Curso de Graduação em Gestão
da Informação

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



Documento assinado eletronicamente por **Jean Carlos Domingos, Coordenador(a)**, em 18/11/2019, às 16:23, 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 18/11/2019, às 17:06, 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 **1660460** e o código CRC **061ED619**.

Referência: Processo nº 23117.094672/2019-84

SEI nº 1660460