



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: 60 horas	CH TOTAL PRÁTICA: 0	CH TOTAL: 60 horas

1. OBJETIVOS

Os objetivos desta disciplina são:

- 1 - Introduzir o aluno às principais tarefas e técnicas de Mineração de Dados.
- 2 – Habilitar o aluno a aplicar ferramentas de Mineração de Dados em problemas práticos.
- 3 – 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 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

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 Arvore de Decisão

4.3 Classificadores baseados nos vizinhos mais próximos (KNN)

4.4 Classificadores baseados em Redes Bayesianas de Crença

4.5 Classificadores baseados em Redes Neurais

4.6 Avaliação de Performance: Método Holdout, Cross-Validation, Bootstrap

4.7 Técnicas de classificação para classes não-balanceadas: métricas alternativas, precisão, recall, Curva Roc.

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, fuzzy, 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

5.5. Método DBSCAN - 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 determinara 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.

9. Aplicações: utilização de ferramentas de Mineração de Dados.

4. BIBLIOGRAFIA BÁSICA

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. 2. ed. Amsterdam ; Boston ; San Francisco: Morgan Kaufmann, 2006.

KANTARDZIC, M. **Data mining: concepts, models, methods, and algorithms**. Hoboken ; Piscataway : John Wiley & Sons : IEEE Press, 2011.

LAROSE, D. T. **Data mining methods and models**. Hoboken; Piscataway : John Wiley & Sons ; IEEE Press, 2006.

5. BIBLIOGRAFIA COMPLEMENTAR

RUSSELL, M.A. **Mineração de dados da Web social**. São Paulo: Novatec, 2011.

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

WITTEN,I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. Amsterdam: Morgan Kaufmann, 2005.

6. APROVAÇÃO



Documento assinado eletronicamente por **Maurício Cunha Escarpinati, Diretor(a)**, em 27/05/2022, às 16:45, 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 **José Fausto de Moraes, Coordenador(a)**, em 28/05/2022, às 16:08, 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 **2476958** e o código CRC **BDC75077**.