



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FICHA DE COMPONENTE CURRICULAR

CÓDIGO: GSI556	COMPONENTE CURRICULAR: MINERAÇÃO DE DADOS	
UNIDADE ACADÊMICA OFERTANTE: FACULDADE DE COMPUTAÇÃO		SIGLA: FACOM
CH TOTAL TEÓRICA: 60	CH TOTAL PRÁTICA: 00	CH TOTAL: 60

OBJETIVOS

Os objetivos da 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.

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.

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 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
9. Aplicações: utilização de ferramentas de Mineração de Dados

BIBLIOGRAFIA BÁSICA

BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer. 2007.

HAN, J.; KAMBER, M.; PEI, J. **Datamining: concepts and techniques**. 3. ed. San Francisco: Morgan Kaufman, 2011.

KUMAR, V.; TAN, P.N.; STEINBACH, M. **Introdução ao Data Mining: mineração de dados**. São Paulo:



Ciência Moderna, 2009.

BIBLIOGRAFIA COMPLEMENTAR

CARVALHO, L. A. V. D. **Datamining**: a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Ciência Moderna, 2005.

NIELSEN, T. D.; JENSEN, F. V. **Bayesian Networks and Decision Graphs**. 2. ed. New York: Springer, 2007.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 4. ed. Boston: Academic Press/Elsevier, 2009.

TORGO, L. **Data mining with R: learning with case studies**. London: Chapman and Hall/CRC, 2010.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Datamining: practical machine learning Tools and techniques**. 3. ed. Boston: Morgan Kaufmann, 2011.

APROVAÇÃO

14 / 03 / 14

2184

Carimbo e assinatura do Coordenador do Curso

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
Prof. Dr. Kil Jin Brandini Park
Coordenador do Curso de Sistema de Informação
Monte Carmelo Portaria R Nº 523/13

14 / 03 / 14

Carimbo e assinatura do Diretor da
Unidade Acadêmica

(que oferece o componente curricular)
Universidade Federal de Uberlândia
Prof. Ilmério Reis da Silva
Diretor da Faculdade de Computação
Portaria R Nº. 757/11