



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
COLEGIADO DO CURSO DE CIÊNCIA DA COMPUTAÇÃO

FICHA DE DISCIPLINA

DISCIPLINA: MINERAÇÃO DE DADOS

CÓDIGO: GBC212		UNIDADE ACADÊMICA: FACULDADE DE COMPUTAÇÃO		
PERÍODO/SÉRIE: -		CH TOTAL TEÓRICA:	CH TOTAL PRÁTICA:	CH TOTAL:
OBRIGATORIA: ()	OPTATIVA: (X)	60	00	60
NÚCLEO DE FORMAÇÃO: Tecnológica / Profissional				
PRÉ-REQUISITOS: NÃO HÁ		CÓ-REQUISITOS: NÃO HÁ		

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.

DESCRIÇÃO DO 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

P.-N. Tan, M. Steinbach, V. Kumar: Introduction to Data Mining, Addison-Wesley, 2005.

J. Han, M. Kamber: Data Mining: Concepts and Techniques, 2a. Ed., Morgan Kaufmann, 2006.

Complementar

I. H. Witten, E. Frank: Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2005.

S. Chakrabarti: Mining the Web: Analysis of Hypertext and Semi Structured Data, Morgan Kaufmann, 2002.

APROVAÇÃO

Carimbo e assinatura do Coordenador do curso
Universidade Federal de Uberlândia
Prof. Ilmério Reis da Silva
Coordenador do Curso de Ciência da Computação
Portaria R nº 713/08

Carimbo e assinatura do Diretor da
Universidade Federal de Uberlândia
Unidade Acadêmica
Prof. Jamil Saadim Beal
Diretor da Faculdade de Computação
Portaria R nº 672/07