

PLANO DE ENSINO

Ficha nº 2

Disciplina: MINERAÇÃO DE DADOS

Código: SIN134

Departamento ministrante: Ciência e Gestão da Informação

Curso: Gestão da Informação

Válido para o ano de: 2010 () 1º semestre (X) 2º semestre

Carga horária semanal: 04 horas Créditos: 03

Aulas: 30 teóricas 30 práticas Total: 60 horas

Turma(s): B Horário(s): terça-feira - 7:30 às 9:30
sexta-feira - 7:30 às 9:30

Pré-requisitos: SIN133 – Banco de Dados

Co-requisitos: SIN116 – Infometria

OBJETIVOS GERAIS DA DISCIPLINA:

Apresentar e aplicar diferentes técnicas estatísticas e computacionais de exploração de dados na identificação de padrões relevantes e não óbvios em bancos de dados.

EMENTA:

Técnicas de descoberta de conhecimento em bancos de dados utilizando as diversas tarefas de mineração de dados.

CONTEÚDO PROGRAMÁTICO:

1. Apresentação (2 horas)
 - 1.1. Posicionamento
 - 1.2. Motivação
 - 1.3. Conceitos Básicos
2. KDD e mineração de dados (6 horas)
 - 2.1. Observando e aprendendo
 - 2.2. Etapas do processo de KDD
 - 2.3. A metodologia de data mining
 - 2.4. As técnicas e tarefas de data mining
 - 2.5. Outros: base de dados; conhecimento; transformação; atributos; partição; substituição; regressão; score; árvore de decisão e relatório.
3. Preparando os dados para data mining (4 horas)
 - 3.1. Dados necessários em data mining
 - 3.2. Estrutura ideal dos dados para data mining
4. Técnicas, heurísticas, algoritmos (2 horas)
 - 4.1. Noções gerais
 - 4.2. Heurísticas diversas
 - 4.2.1. Redes neurais
 - 4.2.2. Lógica nebulosa
 - 4.2.3. Algoritmos genéticos

- 5. Tarefas (12 horas)
 - 5.1. Associação
 - 5.2. Agrupamento
 - 5.3. Classificação

- 6. Pós-processamento (2 horas)
 - 6.1. Simplificações
 - 6.2. Visualização
 - 6.3. Avaliação de resultados

- 7. Aplicações dos algoritmos às bases de dados (30 horas)
 - 7.1. Aplicativo Weka
 - 7.2. Outros aplicativos
 - 7.3. Aplicação dos algoritmos
 - 7.4. Trabalhos

METODOLOGIA:

A disciplina consiste de duas partes distintas:

- 1- Informação Teórica: teoria e algoritmos de descoberta de conhecimento em banco de dados. Itens 1 a 6;
- 2- Análise de algoritmos em laboratório: utilização e interpretação da utilização de algoritmos de mineração de dados em laboratório de informática. Item 7.

A primeira parte será apresentada de forma expositiva, seguidas de tarefas / exercícios de aplicação (conceituais) que devem ser entregues ao final da aula ou no início da aula seguinte (conforme especificado).

A segunda parte, a partir do item 7 será composta de seminários a serem preparados e apresentados pelos próprios alunos sob supervisão dos responsáveis pela disciplina.

FORMAS DE AVALIAÇÃO:

Duas notas principais para composição da média final (média aritmética das duas notas principais), sendo:

Primeira nota: média aritmética da: média das notas das provas escritas (**individuais**), sobre o conteúdo da parte 1 e a média dos exercícios propostos em aula e feitos individualmente ou em equipe (**conforme solicitado**);

Segunda nota: trabalho em grupo sobre um dos temas disponíveis na segunda parte da disciplina (item 7), devendo a avaliação considerar:

- Profundidade na abordagem do tema
- Qualidade do levantamento bibliográfico
- Contextualização do tema na área de interesse
- Explicação do funcionamento do algoritmo em questão
- Exemplificações práticas e contextualizadas do tema
- Análises dos resultados obtidos
- Criatividade da apresentação
- Uso de normas para elaboração de trabalhos de pesquisa (normas ABNT)
- Desempenho individual e coletivo dos membros da equipe durante a apresentação



BIBLIOGRAFIA RECOMENDADA:

Básica

BRAMER, M. **Principles of Data Mining**. Disponível em: <
<http://www.springerlink.com/content/n80326/>>. Acesso em 01 jul. 2010.

CARVALHO, L.A.V. **Datamining** – a mineração de dados no marketing, medicina, economia, engenharia e administração. São Paulo: Editora Moderna, 2005.

LAROSE, D.T. **Data Mining Methods and Models**. Disponível em: <
<http://www3.interscience.wiley.com/cgi-bin/bookhome/110527837>>. Acesso em 01 jul. 2010.

LAROSE, D.T. **Discovering Knowledge in Data**. Disponível em: <
<http://www3.interscience.wiley.com/cgi-bin/bookhome/109880475>>. Acesso em 01 jul. 2010.

MARKOV, Z; LAROSE, D.T. **Data Mining the Web**. Disponível em: <
<http://www3.interscience.wiley.com/cgi-bin/bookhome/112705912>>. Acesso em 01 jul. 2010.

Complementares

AMARAL, F.C.N. **Data Mining: Técnicas e Aplicações para o Marketing Direto**. São Paulo: Editora Berkeley, 2001.

BUSSAB, W.O. , MORETTIN, P.A. **Estatística Básica**. 5.ed. São Paulo: Editora Saraiva, 2002.

BERRY, M.J.A., LINOFF, G. **Data Mining Techniques for Marketing, Sales and Customer Support**. 2ª. ed. New York: John Wiley & Sons, Inc., 2004.

FERNADEZ, G. **Data Mining Using SAS Applications**. New York: Editora Chapman & Hall/CRC, 2003.

GOLDSCHMIDT, R. R. ; PASSOS, E. P. L. **Data Mining: um guia prático - conceitos, técnicas, ferramentas, orientações e aplicações**. 1. ed. Rio de Janeiro: Editora Campus, 2005. v. 1. 250 p.

HAN, J. , KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2001.

KLÖSGEN, W. , ZYTKOW, J.M. **Handbook of Data Mining and Knowledge Discovery**. New York: Oxford University Press, Inc., 2002.

MANLY, B.F.J. **Multivariate Statistical Methods: a primer**. 3.ed. New York: Chapman & Hall, 2005.

TAN, P. N., Steinbach, M. e Kumar, V. **Introduction to Data Mining**. Addison Wesley, 2005.

WEKA. **Waikato Environment for Knowledge Analysis**. The University of Waikato. Disponível em: <
<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 05 jul. 2009.

ASSINATURAS:

Professora Responsável

Denise Fukumi Tsunoda

Decano do colegiado de curso

Leilah Santiago Bufrem

Chefe do Departamento

Mauro José Belli

Prof. Dr. Mauro José Belli
Chefe do Departamento de Ciência e
Gestão da Informação - UFPR
Matr. 120294