

## PLANO DE ENSINO

Ficha nº 2

---

Disciplina: Tópicos em Informática Aplicada II

Código: SIN090

Departamento ministrante: Ciência e Gestão da Informação

Curso: Gestão da Informação

Válido para o ano de: 2008

( ) 1º semestre

( X ) 2º semestre

Carga horária semanal: 04 horas

Créditos:

Aulas: 30 teóricas 30 práticas

Total: 60 horas

Turma(s): A

Horário(s): Terça-feira: 10:30 – 12:30

Sexta-feira: 8:30 – 10:30

### **OBJETIVOS GERAIS DA DISCIPLINA:**

Apresentar e aplicar diferentes técnicas estatísticas e computacionais de exploração de dados para identificar padrões importantes e não óbvios em grandes bancos de dados

### **EMENTA:**

Uso de ferramentas e aplicativos da Informática para gestão da informação.

### **CONTEÚDO PROGRAMÁTICO:**

Parte 0. Breve revisão sobre algoritmos (2 horas)

Parte 1. Introdução (4 horas)

- 1.1 O que é data mining?
- 1.2 Observando e aprendendo
- 1.3 O processo KDD
- 1.4 A metodologia de data mining
- 1.5 As técnicas e tarefas de data mining
- 1.6 Data mining

Parte 2. Exemplo de mineração de dados (2 horas)

- 2.1 Base de dados; Conhecimento; Transformação; Atributos; Partição; Substituição; Regressão; Score; Árvore de decisão e Relatório.

Parte 3. Preparando os dados para data mining (2 horas)

- 3.1 Dados necessários em data mining
- 3.2 Estrutura ideal dos dados para data mining
- 3.3 Entendendo a escala de medidas das variáveis
- 3.4 Amostragem para data mining

Parte 4. Análise exploratória de dados (4 horas)

- 4.1 Distribuições de frequências
- 4.2 Medidas descritivas
- 4.3 Diagramas em caixas
- 4.4 Gráfico normal de probabilidades
- 4.5 Erro padrão e intervalos de confiança
- 4.6 Transformações

#### 4.7 Análise de associação

#### Parte 5. Métodos de aprendizagem não supervisionadas (2 horas)

##### 5.1 Introdução

##### 5.2 Análise de conglomerados

#### Parte 6. Métodos de aprendizagem supervisionados: Predição (6 horas)

##### 6.1 Introdução

##### 6.2 Redes neurais artificiais

#### Parte 7. Métodos de aprendizagem supervisionados: Classificação (6 horas)

##### 7.1 Introdução

##### 7.2 Árvores de decisão

#### Parte 8. Aplicações dos algoritmos às bases de dados (30 horas)

##### 8.1 Aplicativo Weka

##### 8.2 Outros aplicativos

##### 8.3 Aplicação dos algoritmos

##### 8.4 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 7;
- 2- Análise de algoritmos em laboratório: utilização e interpretação da utilização de algoritmos de mineração de dados em laboratório de informática. Item 9.

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

A segunda parte, a partir do item 9 será composta de seminários a serem preparados e apresentados pelos próprios alunos sob supervisão do professor.

**A presença na disciplina é obrigatória, conforme as normas da UFPR;**

### **FORMAS DE AVALIAÇÃO:**

Duas notas principais com peso 5(cinco) cada uma na média final, sendo:

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

Segunda nota: Será a avaliação da apresentação (seminários) em grupo (e por aluno) de um dos temas disponíveis na segunda parte da disciplina (item 9), 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
- Desempenho individual e coletivo dos membros da equipe durante a apresentação

## **BIBLIOGRAFIA:**

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.

BUSSAB, W.O., MIAZAKI, É.S. ANDRADE, D.F. *Introdução à Análise de Agrupamentos*. São Paulo: 9º Simpósio Nacional de Probabilidade e Estatística, 1990.

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

CARVALHO, L.A.V. *Datamining A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. São Paulo: Editora Érica, 2001.

DINIZ, C.A.R., NETOF, L. *Data Mining: Uma Introdução*. São Paulo: XIV Simpósio Nacional de Probabilidade e Estatística. IME-USP, 2000.

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

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.

WEKA. *Waikato Environment for Knowledge Analysis*. The University of Waikato. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 10 jan. 2007.

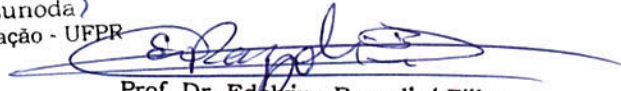
## **ASSINATURAS:**

Professora Responsável

  
Denise Fukumitsu Tsunoda  
Prof. Dr. Denise Fukumitsu Tsunoda  
Depto. de Ciência e Gestão da Informação - UFPR  
Matricula 136.018

Coordenador do Curso

Edelvino Razzolini Filho

  
Prof. Dr. Edelvino Razzolini Filho  
Depto. de Ciência e Gestão da Informação - UFPR  
Matricula 200.114

Chefe do Departamento

  
Helena de Fátima Nunes Silva  
Prof.ª. Dr.ª. Helena de Fátima Nunes Silva  
Chefe do Departamento de Ciência e  
Gestão da Informação - UFPR  
Matr. 103349