

ELE 2760 – Tópicos Especiais em Inteligência Computacional I

Esta disciplina introduz a teoria da Descoberta de Conhecimento em base de dados, apresentando seus conceitos, etapas, principais tarefas e modelos envolvidos, e a aplicação em problemas reais de planejamento financeiro, gerência de produção, diagnóstico, previsão e detecção de fraude, etc. A Descoberta do Conhecimento é uma sub-área da inteligência computacional que trata de descobrir relacionamentos e padrões em grandes massas de dados e texto. A descoberta do conhecimento envolve tecnologias como a mineração dos dados, indução de regras, agrupamento de dados, e outras técnicas relacionadas. Softwares e algumas técnicas e modelos são aplicados em problemas reais de planejamento financeiro, gerência de produção, diagnóstico, previsão, detecção de fraude, etc.

Objetivos: Introduzir a teoria das Data Mining — conceitos básicos e principais modelos de mineração supervisionado e não-supervisionado, fornecendo subsídios para que o aluno saiba discernir quando se deve utilizar as diversas técnicas relacionadas a extração do conhecimento; apresentar ferramentas de software de Data Mining; e exemplificar sua aplicação em sistemas de classificação e reconhecimento de padrões, clusterização, inferência e apoio à decisão.

Ementa: Descoberta de Conhecimento; Conceitos básicos de Bases de Dados e Data Warehouse; Mineração de Dados (MD); 1R, Naive, Árvores de Decisão; Modelos Baseados em Regras, Algoritmos ID3, J48, IBk, EM, Softwares e Ferramentas de MD; Aplicações e Estudos de Casos.

Pré-requisitos: Linguagem de programação Matlab desejável

Professor: Karla Figueiredo
Sala: 4º andar
E-mail: karla@ele.puc-rio.br

Avaliação: *Mestrado/Doutorado:* Prova (P), Trabalho + Artigo Técnico + Apresentação (T2), Exerc. (E)
Critério de Aprovação: Média_{mestrado} = $(3*P + 2*T2 + E1+E2)/7$
Média_{doutorado} = $(3*P + 3*T2 + E1+E2)/7$

Site: <http://www.ica.ele.puc-rio.br> – Este site possui informações complementares sobre o curso e inclui uma **biblioteca com consultas on-line** a livros e periódicos. Também se encontra disponível neste site o documento denominado **ICABook (em cursos)**, contendo: definições de termos na área, ementas dos cursos, listas de exercícios, dicas para uma apresentação oral e dicas para a produção do artigo técnico. Por fim, o site disponibiliza uma base de aplicações (base de casos do ICA) com a descrição sobre aplicações de técnicas inteligentes nos mais diversos problemas.

Bibliografia:

1. Notas de Aula (cópias das transparências – <http://www.ica.ele.puc-rio.br>).
2. WITTEN, Ian H.; Frank, Eibe: "Data Mining", Elsevier
3. HAN, J. , KAMBER, M.. *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2001.
4. KLÖSGEN, W. , ZYTKOW, J.M.. *Handbook of Data Mining and Knowledge Discovery*. New York: Oxford University Press, Inc., 2002.
5. DINIZ, C.A.R. , NETOF, L. *Data Mining: Uma Introdução*. São Paulo: XIV Simpósio Nacional de Probabilidade e Estatística. IME-USP, 2000.
6. BREIMAN Leo, Friedman, Jerome H., Olshen, Richard A., Stone, Charles J. (1984). *Classification and Regression Trees*. Wadsworth International Group, Belmont, California.
7. QUINLAN, Ross (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.
8. HOLTE, R.C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*. 11:63-91.
9. AMARAL, F.C.N. *Data Mining: Técnicas e Aplicações para o Marketing Direto*. São Paulo: Editora Berkeley, 2001.
10. BUSSAB, W.O. , MIAZAKI, É.S. ANDRADE, D.F. *Introdução à Análise de Agrupamentos*. São Paulo: 9º Simpósio Nacional de Probabilidade e Estatística, 1990.
11. BERRY, M.J.A., LINOFF, G. *Data Mining Techniques For Marketing, Sales and Customer Support*. 2a. ed. New York: John Wiley & Sons, Inc., 2004.

12. CARVALHO, L.A.V. *Datamining A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. São Paulo: Editora Érica, 2001.
13. FERNANDEZ, G. *Data Mining Using SAS Applications*. New York: Editora Chapman & Hall/CRC, 2003.
14. MANLY, B.F.J. *Multivariate Statistical Methods: a Primer*. 3.ed. New York: Chapman & Hall, 2005.
15. OLIVEIRA, I. (2003). *Correlated Data in Multivariate Analysis. Ph.D Thesis*, University of Aberdeen.
16. MCCULLAGH, P. and Nelder, J.A., *Generalized Linear Models* 2nd edition, Chapman & Hall 1997
17. BUSSAB, W.O. , MORETTIN, P.A. *Estatística Básica*. 5.ed. São Paulo: Editora Saraiva, 2002.

CRONOGRAMA DO CURSO

Etapas	Datas
<i>Entrega 1º Exercício – Tratamento /Seleção de de Variáveis</i>	12/05
<i>Definição e apresentação dos Temas dos Projetos</i>	26/05
<i>Entrega 2º Exercício – Classificação/Inferência de Dados</i>	09/06
<i>Entrega do artigo (Preliminar)</i>	30/06
<i>Prova</i>	30/06
<i>Final dos Projetos (Apresentação+Artigo+CD)</i>	07/07