

MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA CELSO SUCKOW DA FONSECA
DIRETORIA DE ENSINO (DIREN)
DEPARTAMENTO DE ENSINO SUPERIOR (DEPES)
DEPARTAMENTO DE INFORMÁTICA (DEPIN)
CURSO SUPERIOR DE TECNOLOGIA EM SISTEMAS PARA INTERNET (CST-SI)

DEPARTAMENTO/ COORDENAÇÃO
DEPIN - Departamento Acadêmico de Informática

PLANO DE CURSO DA DISCIPLINA
MINERAÇÃO DE DADOS

CÓDIGO DA DISCIPLINA
GTSI1298

PERÍODO
Opt

ANO
2012

SEMESTRE
2

PRÉ-REQUISITOS
GTSI1245 PROJETO DE BANCO DE DADOS

CRÉDITOS
4

AULAS/SEMANA		
TEÓRICA	PRÁTICA	ESTÁGIO
4	0	0

TOTAL DE AULAS NO SEMESTRE
72

EMENTA

Estudo de técnicas de mineração de dados, i.e., extração de conhecimento a partir de grandes volumes de dados. O processo de extração de conhecimento contempla a análise exploratória de dados, pré-processamento, identificação de outliers, padrões frequentes, classificação, agrupamento e *datawarehouses*.

BIBLIOGRAFIA

Bibliografia básica:

- HAN, M. KAMBER e PEI J. Data Mining: Concepts and Techniques, Editora Morgan Kaufmann, Burlington, MA, EUA.
- ZAKI, M.J. and Jr. W.M. Data Mining and Analysis: Fundamental Concepts and Algorithms, Editora Cambridge University Press, Cambridge, United Kingdom.
- WITTEN, I.H., FRANK, E. e HALL M.A., Data Mining: Practical Machine Learning Tools and Techniques, Editora Morgan Kaufmann, Burlington, MA, EUA.

Bibliografia complementar:

- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Editora Springer, New York, EUA.
- JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R., An Introduction to Statistical Learning: with Applications in R. , Editora Springer, New York, EUA.
- LANTZ, B., Machine Learning with R. Editora Packt Publishing, United Kingdom..
- LESKOVEC, J., RAJARAMAN, A., ULLMAN, J. D., Mining of Massive Datasets. Editora Cambridge University Press, Cambridge, United Kingdom.
- SHUMWAY, R. H., STOFFER, D. S., Time Series Analysis and Its Applications: With R Examples. Editora Springer, New York, EUA.

OBJETIVO GERAL

Dominar o processo de mineração de dados e todos os elementos relacionados ao processo de extração de conhecimento a partir de grandes volumes de dados. É dada uma ênfase importante no pré-processamento de dados, uma vez que o conhecimento dos dados auxilia na identificação das mais adequadas técnicas de mineração de dados a serem adotadas.

METODOLOGIA

- Aulas expositivas com recursos audiovisuais
- Desenvolvimento de estudos de casos em laboratório, com uso de desenvolvimento dirigido por testes.
- Atendimento e orientação do professor visando o desenvolvimento de uma monografia.

CRITÉRIOS DE AVALIAÇÃO

A avaliação semestral envolve uma prova (P) e uma monografia (M). As datas das provas são agendadas entre o professor e a turma. A média parcial (MP) será calculada pelo cômputo da média ponderada simples entre as notas dos seminários e monografia:

$$MP = 0.5 * P + 0.5 * M$$

Segundo o regimento do CEFET-RJ, caso o aluno obtenha média parcial inferior a 3,0 (três e zero) estará reprovado diretamente. Graus MP maiores ou iguais a 7,0 (sete e zero) aprovam diretamente o aluno. Em situações onde o aluno tenha grau MP entre 3,0 inclusive e 7,0 exclusive, terá direito a uma prova final (PF), que, juntamente com a média parcial gerará uma nova média, denominada média final (MF). Essa média é calculada da seguinte forma:

$$MF = (MP + PF) / 2$$

Para ser aprovado, o aluno deve alcançar uma média final MF maior ou igual a 5,0 (cinco e zero). Caso contrário, estará reprovado, devendo repetir a componente curricular.

PROGRAMA

1. Introdução
 - a. Definição de mineração de dados
 - b. Aplicações de mineração de dados
 - c. Desafios em mineração de dados
2. Entendendo melhor os dados
 - a. Tipos de objetos e atributos
 - b. Estatística básica
 - c. Visualização de dados
 - d. Medições de similaridade/dissimilaridade
3. Preparação (pré-processamento) de dados
 - a. Limpeza de dados
 - b. Integração de dados
 - c. Redução de dados
 - d. Análise de componentes principais
 - e. Identificação de outliers
 - f. Transformação/Discretização
4. Classificação
 - a. Lazy Learning
 - b. Naive Bayes
 - c. Árvore de regressão
 - d. Redes Neurais
 - e. SVM
5. Agrupamento
 - a. Métodos de partição

- b. Métodos hierárquicos
 - c. Densidade
- 6. Padrões frequentes
- 7. Datawarehouse
 - a. Introdução
 - b. Data Cube/OLAP