

Programa de Ensino

1) Identificação

Disciplina: INE5644 - Data Mining
Carga horária: 72 horas-aula Teóricas: 64 Práticas: 8
Período: 1º semestre de 2020 até a presente data

2) Cursos

- Ciências da Computação (208)
- Engenharia, áreas Elétrica e Mecânica, habilitação Controle e Automação (220)
- Sistemas de Informação (238)

3) Requisitos

- Ciências da Computação (208)
 - INE5423 - Banco de Dados I
- Sistemas de Informação (238)
 - INE5606 - Probabilidade e Estatística
 - INE5633 - Sistemas Inteligentes
 - INE5643 - Data Warehouse

4) Ementa

Considerações iniciais da área e aplicações. Processos de descoberta do conhecimento (KDD) em base de dados. Análise exploratória de dados. Tipos de variáveis, Técnicas de Data Mining para classificação, estimação, predição, análise de agrupamentos, análise de associação: redes neurais, árvores de decisão, regras de decisão, análise discriminante, regressão linear, regressão logística, análise de cluster, análise de componentes principais. Uso de softwares de data mining. Exercícios. Data Mining de caso.

5) Objetivos

Geral: Apresentar e aplicar as principais técnicas de mineração de dados, o processo de KDD com vistas à identificação de padrões importantes e não óbvios em grandes bancos de dados.

Específicos:

- Assimilar os principais conceitos acerca de Mineração de Dados (MD);
- Compreender os passos do processo de descoberta de conhecimento em base de dados (KDD);
- Realizar a análise exploratória dos dados;
- Compreender as principais tarefas da mineração de dados e relacionar as respectivas técnicas;
- Conhecer e saber aplicar as principais técnicas de mineração de dados (classificação, agrupamentos, estimação, predição e associação);
- Utilizar ferramenta para uso de algoritmos de mineração de dados.

6) Conteúdo Programático

- 6.1) Introdução e Conceitos [6 horas-aula]
- O que é KDD

- Tarefas, processos e técnicas
- 6.2) Processo KDD [6 horas-aula]
 - As etapas do processo KDD
- 6.3) Análise Estatística de Dados [8 horas-aula]
 - Estatística descritiva dos dados
 - Normalização de dados
 - Transformação de dados
- 6.4) Classificação [12 horas-aula]
 - Abordagem simbólica
 - Algoritmo ID3
 - Algoritmo C45
 - Abordagem estatística
 - Algoritmo de Naive-Bayes
 - Abordagem conexionista
 - Redes Neurais artificiais (MLP e RBF)
 - Uso de ferramenta
 - Trabalho prático
- 6.5) Agrupamento [10 horas-aula]
 - Algoritmos de clustering convencionais (Kmeans)
 - Algoritmos hierárquicos (Dendogramas)
 - Trabalho prático
- 6.6) Associação [10 horas-aula]
 - Algoritmo apriori (regras de associação)
 - Trabalho prático
- 6.7) Tópicos Avançados [10 horas-aula]
 - Seminários sobre tópicos atuais
- 6.8) Projeto [10 horas-aula]

7) Bibliografia Básica

- WITTEN, I. H. et al. Data Mining: Practical Machine Learning Tools and Techniques. 4a Edição. Amsterdam: Morgan Kaufmann, 2017.
- IGUAL, L. et al. Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications. 1.ed. Springer Publishing Company, Incorporated, 2017
- SKIENA, S. S. The Data Science Design Manual. 1.ed. Springer Publishing Company, Incorporated, 2017.
- BRAMER, M. Principles of Data Mining. 3.ed. Springer Publishing Company, Incorporated, 2016.
- AGGARWAL, C. C. Data Mining: The Textbook. Springer Publishing Company, Incorporated, 2015

8) Bibliografia Complementar

- TAN, P. et al. Introduction to Data Mining. 2.ed. Pearson, 2018.
- 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.
- CARVALHO, L.A.V. Data Mining 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. New York: Editora Morgan Kaufmann Publisher, 2001.
- MANLY, B.F.J. Multivariate Statistical Methods: A primer. 3a.ed. New York: Chapman & Hall, 2005.