

IND 2622

Ciência de Dados para Processos de Negócio

CARGA HORÁRIA TOTAL: 45

CRÉDITOS: 03

CRITÉRIO: 12

PRÉ-REQUISITO(S): -----

OBJETIVOS

Capacitar o aluno para a condução de um projeto de Ciência de Dados para apoio a Processos de Negócio, conhecendo conceitos, técnicas e algoritmos de pré-processamento, técnicas de mineração de dados e pós-processamento, experimentando a aplicação destas técnicas no cenário de um processo de negócio. Apresentar ao aluno aplicações de Ciência de Dados sobre dados não estruturados. Apresentar ao aluno desafios de pesquisa e cenários de aplicação recentes de Ciência de Dados, em especial como apoio a Processos de Negócio Intensivos em Conhecimento.

EMENTA

Introdução e exemplos de aplicações da Ciência de Dados em contextos de Processos de Negócio. Conceitos e exemplos de Processos de Negócio e Processos Intensivos em Conhecimento. Ciclo de vida da Ciência de Dados. Noções de metodologia científica experimental. Processo de descoberta de conhecimento em bases de dados: Pré-processamento, Técnicas de Mineração de Dados (Classificação, Agrupamento, Regras de associação e outras selecionadas no estado da arte), Pós-processamento. Noções de mineração de dados não estruturados. Noções de Mineração de Processos. Desafios de pesquisa.

PROGRAMA

- Introdução
 - Motivação e aplicações da Ciência de Dados em contextos de Processos de Negócio
 - Conceitos de Processos de Negócio e Processos Intensivos em Conhecimento
 - Panorama das principais técnicas de Ciência de Dados
 - Aprendizado supervisionado e não-supervisionado
- Ciclo de vida da Ciência de Dados
 - Entendimento do Problema
 - Coleta de Dados
 - Ciclo interno de Pesquisa
 - Processo de descoberta do conhecimento em base de dados
 - Visualização de Resultados
 - Criação de Ações com base nos resultados
 - Coleta de Feedback das Ações
- Noções de Metodologia científica experimental
 - Experimento científico
 - Exemplos de Sistemas de Gerência de Workflows Científicos
- Processo de descoberta do conhecimento em base de dados

- Pré-processamento
 - Tipos de Dados: Convencional, Relacional, Espacial, Temporal, Textual; Estruturas e Modelos de Representação; Padrões Preditivos e Descritivos;
 - Técnicas: Normalização, tratamento de dados faltantes, redução de dimensionalidade, seleção de observações e amostragem, tratamento de ruídos, Engenharia de características
- Pós-processamento
 - Métricas para avaliação de modelos
 - Metodologia experimental para avaliação de modelos
- Técnicas de mineração de dados (estruturados)
 - Classificação
 - Aprendizado de regras
 - Árvores de decisão
 - Agrupamento (Clustering)
 - Redes Neurais Artificiais
 - Regras de associação
 - Outras técnicas selecionadas no estado da arte
- Noções de Mineração de dados não estruturados
 - Conceitos e aplicações
 - Mineração de Texto e Processamento em Linguagem Natural
 - Mineração de Processos
 - Mineração de Processos no ciclo de vida BPM
- Desafios de pesquisa
 - Seleção, Leitura e análise de artigos científicos sobre o estado da arte no tema

AVALIAÇÃO

Avaliação contínua: Ao longo da disciplina, o aluno entregará exercícios e tarefas intermediárias, aplicando e fixando individualmente o conteúdo da disciplina. Ao final da disciplina, o aluno desenvolverá um projeto de Ciência de Dados aplicado a um Processo de Negócio específico, a ser definido e acordado com o professor ao longo da disciplina.

Avaliação Somativa: Notas dos exercícios e tarefas intermediárias, e nota do projeto.

BIBLIOGRAFIA PRINCIPAL

- Han J, Kamber M, Pei J, Data Mining: Concepts and Techniques. 3a edição. Morgan Kaufmann, 2011
- Zaki M, Meira W: Data Mining and Machine Learning: Fundamental Concepts and Algorithms. 2a edição. Cambridge University Press, 2020.
- Dumas M, La Rosa M, Mendling J, Reijers HA. Fundamentals of Business process management. 2a edição. Springer-Verlag; 2018.

BIBLIOGRAFIA COMPLEMENTAR

- Aalst W. Process mining – Data science in action. 2a edição. Springer; 2016.
- Feldman e Sanger, The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data, Cambridge University Press, 2007
- Witten, Frank, Hall e Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4 Edition, Ed. Morgan Kaufmann, 2016
- Wiley & Sons, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services (Ed.), 2015
- Shcherbakov, Maxim, Shcherbakova, Nataliya, Brebels, Adriaan, Janovsky, Timur, Kamaev, Valery, 2014, "Lean Data Science Research Life Cycle: A Concept for Data Analysis Software Development", Knowledge-Based Software Engineering, v. 466, pp. 708-716, Springer International Publishing, 2014.
- Artigos científicos selecionados ao longo da disciplina