

PROGRAMA DE DISCIPLINA

DEPARTAMENTO: Tecnologia Industrial

DISCIPLINA: CIÊNCIA DE DADOS

SIGLA: 8CID004

CARGA HORÁRIA TOTAL: 36h

*

TEORIA:

36h*

PRÁTICA:

00h*

EXTENSÃO: 00h*

CURSO: Bacharelado em Sistemas de Informação

PRÉ-REQUISITOS: 3DAD104

EMENTA: Introdução a Ciência dos Dados. Técnicas e ferramentas para Ciência de Dados. Aprendizado e análise exploratória de dados.

OBJETIVO GERAL: Capacitar o aluno a compreender e executar o ciclo de vida da Ciência de Dados, integrando o pensamento analítico voltado a negócios e descobertas científicas com práticas modernas de manipulação, modelagem estatística e visualização de informações.

OBJETIVOS ESPECÍFICOS:

Contextualizar o impacto do Quarto Paradigma científico e estruturar problemas reais sob a ótica da Ciência de Dados;

Dominar o uso de linguagens de programação e bibliotecas essenciais para coleta, limpeza e manipulação de dados;

Aplicar conceitos fundamentais de mineração de dados e aprendizado estatístico para criar e avaliar modelos preditivos e descritivos básicos;

Projetar visualizações de dados eficientes, tanto para a análise exploratória quanto para a comunicação clara dos resultados aos tomadores de decisão.

PROGRAMA

1. Introdução à Disciplina
 - 1.1. Apresentação da disciplina
 - 1.2. Metodologia de ensino utilizada
 - 1.3. Avaliação
2. Fundamentos da Ciência de Dados e Pensamento Analítico
 - 2.1. O Quarto Paradigma e a descoberta científica intensiva em dados
 - 2.2. Ciência de dados aplicada a negócios e formulação de problemas
 - 2.3. O ciclo de vida de um projeto de dados
3. Programação e Manipulação de Dados
 - 3.1. Introdução ao ecossistema de programação para dados
 - 3.2. Estruturas de dados matemáticas e tabulares
 - 3.3. Técnicas de coleta, limpeza, filtragem e transformação de dados
4. Aprendizado Estatístico e Modelagem Preditiva
 - 4.1. Fundamentos de aprendizado de máquina, viés, variância e validação
 - 4.2. Algoritmos supervisionados principais
 - 4.3. Algoritmos não supervisionados

Bibliografia Básica

HEY, T., TANSLEY, S. and TOLLE, K.M., 2009. The fourth paradigm: data-intensive scientific discovery (Vol. 1). Redmond, WA: Microsoft research.

MUNZNER, T., 2014. Visualization Analysis and Design. CRC Press.

PROVOST, F. and FAWCETT, T., 2013. Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc.

Bibliografia Complementar

JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R., 2013. An introduction to statistical learning (Vol. 6). New York: Springer.

ZAKI, M.J. and MEIRA Jr, W., 2014. Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.

MCKINNEY, W., 2012. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.

GRUS, J., 2015. Data Science from Scratch: First Principles with Python. O'Reilly Media, Inc.
WICKHAM, H., 2009. ggplot2: elegant graphics for data analysis. Springer Science & Business Media.