



Plano de Ensino

CAMPUS VARGINHA

DISCIPLINA: Ciência de Dados

CÓDIGO: G08CIDA0.01

Início: **02/2026**

Carga Horária: Total: 30 horas/aula Semanal: 02 aulas/aula Créditos: 02

Natureza: Teórica / Obrigatória

Área de Formação - DCN: Profissionalizante

Departamento que oferta a disciplina: Departamento de Computação e Engenharia Civil

Ementa:

Alteração de dados. Gráficos e visualização. Agregação de dados e operação de grupos. Álgebra Linear. Probabilidade. Modelos como k-vizinhos mais próximos. Naive Bayes. Árvores de decisão. Agrupamentos. Séries temporais. Aplicação de dados econômicos e financeiros. Processamento de linguagem natural. Análise de rede. MapReduce e bases de dados.

Curso(s)	Período	Eixo	Obrigatória	Optativa
Bacharelado em Sistemas de Informação	8º	Engenharia de Dados e Informação	X	

INTERDISCIPLINARIDADES

Prerrequisitos

Mineração de Dados

Laboratório de Mineração de Dados

Análise Multivariada

Correquisitos

Laboratório de Ciência de Dados

Objetivos: *A disciplina deverá possibilitar ao estudante*

1	Identificar conceitos-chave relacionados à alteração de dados, gráficos e visualização, agregação de dados, e operação de grupos.
2	Explicar os princípios subjacentes à criação e interpretação de gráficos e visualizações.
3	Utilizar modelos como k-vizinhos mais próximos, Naive Bayes, árvores de decisão e agrupamentos para resolver problemas práticos.
4	Aplicar conceitos de séries temporais na análise de dados econômicos e financeiros.
5	Implementar processamento de linguagem natural em projetos específicos.
6	Avaliar o impacto e a relevância das técnicas de MapReduce na manipulação eficiente de grandes conjuntos de dados.

Plano de Ensino

Unidades de ensino	Carga-horária Horas/aula
1 Introdução 1.1 A Ascensão dos Dados 1.2 O Que É Data Science? 1.3 Motivação Hipotética: DataSciencester	01
2 Obtendo Dados 2.1 Stdin e stdout 2.2 Lendo Arquivos 2.3 Extraíndo Dados da Internet 2.4 Usando APIs	01
3 Visualizando Dados 3.1 Matplotlib 3.2 Gráficos de Barra 3.3 Gráficos de Linhas 3.4 Gráficos de Dispersão	01
4 Álgebra Linear 4.1 Vetores 4.2 Matrizes	01
5 Estatística 5.1 Descrevendo um Conjunto Único de Dados 5.2 Correlação 5.3 Paradoxo de Simpson 5.4 Alguns Outros Pontos de Atenção sobre Correlação 5.5 Correlação e Causalidade	01
6 Probabilidade 6.1 Dependência e Independência 6.2 Probabilidade Condicional 6.3 Teorema de Bayes 6.4 Variáveis Aleatórias 6.5 Distribuições Contínuas 6.6 A Distribuição Normal 6.7 O Teorema do Limite Central	01
7 Trabalhando com Dados 7.1 Explorando Seus Dados 7.2 Limpando e Transformando 7.3 Manipulando Dados 7.4 Redimensionando 7.5 Redução da Dimensionalidade	02
8 Aprendizado de Máquina 8.1 Modelagem 8.2 O Que É Aprendizado de Máquina? 8.3 Sobreajuste e Sub-Ajuste 8.4 Precisão 8.5 Compromisso entre Polarização e Variância 8.6 Recursos Extração e Seleção de Característica	04
9 K-Vizinhos Mais Próximos	02

Plano de Ensino

	9.1 O Modelo 9.2 Exemplo: Linguagens Favoritas 9.3 A Maldição da Dimensionalidade	
10	Naive Bayes 10.1 Um filtro de Spam simples 10.2 Um Filtro de Spam Mais Sofisticado 10.3 Implementação 10.4 Testando Nosso Modelo	02
11	Regressão Múltipla 11.1 O Modelo 11.2 Mais Suposições do Modelo dos Mínimos Quadrados 11.3 Ajustando o Modelo 11.4 Interpretando o Modelo 11.5 O Benefício do Ajuste 11.6 Digressão: A Inicialização 11.7 Erros Padrões de Coeficientes de Regressão 11.8 Regularização	02
12	Regressão Logística 12.1 O Problema 12.2 A Função Logística 12.3 Aplicando o Modelo 12.4 O Benefício do Ajuste 12.5 Máquina de Vetor de Suporte	02
13	Árvores de Decisão 13.1 O que é uma Árvore de Decisão? 13.2 Entropia 13.3 A Entropia de uma partição 13.4 Criando uma Árvore de Decisão 13.5 Juntando Tudo 13.6 Florestas aleatórias	02
14	Agrupamento 14.1 A Ideia 14.2 O Modelo 14.3 Exemplo: Encontros 14.4 Escolhendo k 14.5 Exemplo: Agrupando Cores 14.6 Agrupamento Hierárquico Bottom-up	02
15	Processamento de Linguagem Natural 15.1 Nuvens de Palavras 15.2 Modelos n-gramas 15.3 Gramáticas 15.4 Um Adendo: Amostragem de Gibbs 15.5 Modelagem de Tópicos	02
16	Análise de Rede 16.1 Centralidade de Intermediação 16.2 Centralidade de Vetor Próprio 16.3 Gráficos Direcionados e PageRank	02
17	MapReduce	02

Plano de Ensino

17.1 Exemplo: Contagem de Palavras 17.2 Por que MapReduce? 17.3 MapReduce Mais Generalizado 17.4 Exemplo: Analisando Atualizações de Status 17.5 Exemplo: Multiplicação de Matriz 17.6 Um Adendo: Combinadores	
Total	30

Bibliografia Básica

1	GRUS, Joel. Data Science do Zero. Primeiras Regras com o Python . Rio de Janeiro: Alta Books, 2016, 336 p.
2	McKINNEY, Wes. Python for Data Analysis . Sebastopol, CA: O'Reilly Media, 2013. 452 p.
3	PROVOST, F.; FAWCETT, T. Data Science para Negócios - O que você precisa saber sobre mineração de dados e pensamento analítico de dados . Rio de Janeiro: Alta Books, 2016, 408 p.

Bibliografia Complementar

1	BORGES, L.E. Python para Desenvolvedores: aborda Python 3.3 . 1. ed. Novatec, 2014.
2	CAMPOS, Frederico Ferreira. Algoritmos Numéricos . 2. ed. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 2007. 428p.
3	MATTHEUS, E. Curso Intensivo De Python: Uma Introdução Prática E Baseada Em Projetos À Programação . São Paulo: Novatec, 2016.
4	RAMALHO, LUCIANO. Python Fluente: Programação Clara, Concisa e Eficaz . 1. Ed. Novatec, 2015.
5	ZIVIANI, NIVIO. Projeto de Algoritmos: com Implementações em Java e C++ . São Paulo: Thomson Learning, C2007. XX, 621 P., IL. ISBN 85-221-0525-1.