

**CAMPUS VARGINHA**

**DISCIPLINA:** Laboratório de Ciência de Dados

**CÓDIGO:** G08LCDA0.01

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**

Ciência de Dados

**Objetivos:** *A disciplina deverá possibilitar ao estudante*

|   |                                                                                                                                                                       |
|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Aplicar técnicas práticas de agregação de dados e operações em grupos utilizando ferramentas específicas de ciência de dados.                                         |
| 2 | Desenvolver modelos práticos de machine learning, como k-vizinhos mais próximos, Naive Bayes, árvores de decisão e agrupamentos, utilizando conjuntos de dados reais. |
| 3 | Integrar técnicas de processamento de linguagem natural em projetos práticos, como a análise de sentimentos em textos.                                                |
| 4 | Integrar a análise de rede em problemas práticos, utilizando conjuntos de dados reais.                                                                                |
| 5 | Avaliar a performance de modelos de machine learning em cenários práticos, utilizando métricas específicas.                                                           |
| 6 | Avaliar a eficácia de técnicas de MapReduce na manipulação prática de grandes                                                                                         |

Plano de Ensino

|  |                     |
|--|---------------------|
|  | conjuntos de dados. |
|--|---------------------|

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

Plano de Ensino

|    |                                                                                                                                                                                                                                                                                                        |    |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|    | 8.6 Recursos Extração e Seleção de Característica                                                                                                                                                                                                                                                      |    |
| 9  | <b>K-Vizinhos Mais Próximos</b><br>9.1 O Modelo<br>9.2 Exemplo: Linguagens Favoritas<br>9.3 A Maldição da Dimensionalidade                                                                                                                                                                             | 02 |
| 10 | <b>Naive Bayes</b><br>10.1 Um filtro de Spam simples<br>10.2 Um Filtro de Spam Mais Sofisticado<br>10.3 Implementação<br>10.4 Testando Nosso Modelo                                                                                                                                                    | 02 |
| 11 | <b>Regressão Múltipla</b><br>11.1 O Modelo<br>11.2 Mais Suposições do Modelo dos Mínimos Quadrados<br>11.3 Ajustando o Modelo<br>11.4 Interpretando o Modelo<br>11.5 O Benefício do Ajuste<br>11.6 Digressão: A Inicialização<br>11.7 Erros Padrões de Coeficientes de Regressão<br>11.8 Regularização | 02 |
| 12 | <b>Regressão Logística</b><br>12.1 O Problema<br>12.2 A Função Logística<br>12.3 Aplicando o Modelo<br>12.4 O Benefício do Ajuste<br>12.5 Máquina de Vetor de Suporte                                                                                                                                  | 02 |
| 13 | <b>Árvores de Decisão</b><br>13.1 O que é uma Árvore de Decisão?<br>13.2 Entropia<br>13.3 A Entropia de uma partição<br>13.4 Criando uma Árvore de Decisão<br>13.5 Juntando Tudo<br>13.6 Florestas aleatórias                                                                                          | 02 |
| 14 | <b>Agrupamento</b><br>14.1 A Ideia<br>14.2 O Modelo<br>14.3 Exemplo: Encontros<br>14.4 Escolhendo k<br>14.5 Exemplo: Agrupando Cores<br>14.6 Agrupamento Hierárquico Bottom-up                                                                                                                         | 02 |
| 15 | <b>Processamento de Linguagem Natural</b><br>15.1 Nuvens de Palavras<br>15.2 Modelos n-gramas<br>15.3 Gramáticas<br>15.4 Um Adendo: Amostragem de Gibbs<br>15.5 Modelagem de Tópicos                                                                                                                   | 02 |
| 16 | <b>Análise de Rede</b><br>16.1 Centralidade de Intermediação<br>16.2 Centralidade de Vetor Próprio                                                                                                                                                                                                     | 02 |

Plano de Ensino

|       |                                                                                                                                                                                                                                                   |    |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|       | 16.3 Gráficos Direcionados e PageRank                                                                                                                                                                                                             |    |
| 17    | <b>MapReduce</b><br>17.1 Exemplo: Contagem de Palavras<br>17.2 Por que MapReduce?<br>17.3 MapReduce Mais Generalizado<br>17.4 Exemplo: Analisando Atualizações de Status<br>17.5 Exemplo: Multiplicação de Matriz<br>17.6 Um Adendo: Combinadores | 02 |
| Total |                                                                                                                                                                                                                                                   | 30 |

**Bibliografia Básica**

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

**Bibliografia Complementar**

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