



## Plano de Ensino

**Universidade Federal do Espírito Santo**

**Campus de Goiabeiras**

**Curso:** Arquivologia - Noturno

**Departamento Responsável:** Departamento de Arquivologia

**Data de Aprovação (Art. nº 91):**

**DOCENTE PRINCIPAL :** HENRIQUE MONTEIRO CRISTOVAO

Matrícula: 1727965

**Qualificação / link para o Currículo Lattes:** <http://lattes.cnpq.br/5035919384923489>

**Disciplina:** CIÊNCIAS DE DADOS

**Código:** ARV12967

**Período:** 2023 / 1

**Turma:** 01

**Pré-requisito:**

**Carga Horária Semestral:** 60

Disciplina: ARV13049 - SISTEMAS GERENCIADORES DE BANCOS DE DADOS (SGDB)

### Distribuição da Carga Horária Semestral

**Créditos:** 2

**Teórica**

**Exercício**

**Laboratório**

30

15

15

### Ementa:

Metodologias e técnicas da ciência de dados para armazenamento, análise, tratamento, acessibilidade e visualização de dados.

### Objetivos Específicos:

O aluno deverá conhecer as metodologias e técnicas da ciência de dados para armazenamento, análise, tratamento, acessibilidade e visualização de dados.

### Conteúdo Programático:

Unidade 1: Dados

- 1.1 Data Literacy
- 1.2 Data Driven, Datafication e Dataism
- 1.3 Infometria (ou informetria)
- 1.4 Dados Relacionados com Informação e Conhecimento
- 1.5 Dados na Internet e Web de dados
- 1.6 Dados Abertos (Open Data)
- 1.7 Princípios FAIR
- 1.8 LGPD: Anonimização e Pseudonimização de Dados Pessoais
- 1.9 Categorias de Dados e Variáveis
- 1.10 Softwares de Manipulação de Dados

Unidade 2: Introdução à Ciência de Dados

- 2.1 Contexto e Aplicações
- 2.2 Correlação entre Variáveis
- 2.3 Manipulações Tendenciosas
- 2.4 Mineração de Dados (Data Mining)
- 2.5 Descoberta de Conhecimento
- 2.6 Visualização de Dados, Informação e Conhecimento
- 2.7 Data Storytelling
- 2.8 Metodologias e Técnicas para Ciência de Dados

Unidade 3: Introdução a Banco de Dados

- 3.1 Banco de Dados Relacional
- 3.2 Modelo de Dados
- 3.3 Sistema Gerenciador de Banco de Dados (SGBD)
- 3.4 Linguagem de Consulta SQL

Unidade 4: Introdução à Ciência das Redes

- 4.1 Redes Complexas
- 4.2 Redes Sociais
- 4.3 Redes de Informação
- 4.4 Cronologia, Modelos e Marcos Históricos

#### Unidade 5: Análise de Redes de Informação

- 5.1 Fundamentos sobre Análise de Redes
- 5.2 Conceitos Básicos de Redes
- 5.3 Métricas de Centralidade
- 5.4 Coesão, Comunidades e Contágio
- 5.5 Redes Bipartidas (bipartite)
- 5.6 Projeção Bipartida

#### Softwares principais:

- a) OpenRefine
- b) Power BI Desktop
- c) MySQL Workbench
- d) Gephi

#### Estudos de Casos

Prática sobre contextos reais em softwares de limpeza e preparação de dados, recuperação de informação com SGBD, visualização de dados com dashboard, análise de redes por inspeção visual.

#### Metodologia:

Aulas expositivas e dialogadas com auxílio de quadro branco e projetor multimídia.

Estímulo à aprendizagem por meio do desenvolvimento de atividades individuais e coletivas.

Uso da metodologia aprendizagem baseada em problemas (PBL - problem based learning). O contexto a ser explorado por meio do PBL será o desenvolvimento de uma análise exploratória sobre uma base de dados abertos escolhida pelo aluno de forma que ele possa aplicar, ao longo do semestre letivo, os tópicos estudados conforme indicação do professor.

Criação e alimentação de um diário de bordo, isto é, um registro semanal e individual, cronológico e dinâmico que contém resultados e reflexões acerca das atividades desenvolvidas e conteúdo aprendido ao longo da disciplina.

Uso do Ambiente Virtual de Aprendizagem da UFES (AVA: <http://ava.ufes.br>) para apoio às atividades presenciais e a distância.

Algumas atividades da disciplina poderão acontecer a distância, conforme previsto na portaria MEC nº1.134 de 10 de outubro de 2016.

#### Critérios / Processo de avaliação da Aprendizagem :

A avaliação da aprendizagem dos alunos irá considerar os aspectos qualitativos e quantitativos de obtenção da nota final e por meio da promoção de uma pedagogia da inclusão e obedecendo os objetivos e os conteúdos explicitados no presente plano de ensino.

Cada avaliação terá um peso de acordo com a sua importância conforme análise do professor. A soma dos pesos de todas as notas será 10 e cada nota será normalizada em valores de 0 a 1,0. A média parcial será calculada pela soma de todas as avaliações multiplicadas pelo seu respectivo peso, resultando em um valor de 0 a 10.

Se a média parcial for maior ou igual a 7,0 o aluno será aprovado, caso contrário, o aluno terá oportunidade de fazer uma prova final com pontuação de 0 a 10. Nesse caso, a média final será:

(média parcial + nota da prova final) / 2, e a aprovação será obtida com um valor maior ou igual 5.

#### Bibliografia básica:

CARVALHO, L. A. V. – Data Mining: A mineração de dados no marketing, medicina, engenharia e administração. São Paulo: Érica, 2001. FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMITH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In Proceedings of the Second International Conference on Data Mining and Knowledge Discovery, AAAI Press, Menlo Park, US; 1996. Disponível em [kdd96-014.pdf](#). Acesso em Maio de 2016. REZENDE, Solange Oliveira. Mineração de Dados. Anais do V Encontro Nacional de Inteligência Artificial, Sociedade Brasileira de Computação. ISBN: 8576690330. São Leopoldo RS, 25 a 29 de julho de 2005. Disponível em [0102.pdf](#). Acesso em Maio de 2016. SILVA, Leandro. A.; SILVA Luciano. Fundamentos de Mineração de Dados Educacionais. Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 3º Congresso Brasileiro de Informática na Educação (CBIE 2014) Workshops (WCBIE 2014). Disponível em [cbie.wcbie.2014.568](#). Acesso em Maio de 2016.

#### Bibliografia complementar:

BAKER, R.; ISOTANI, S.; DE CARVALHO, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, Volume 19, Número 2, 2011. Disponível em 67 . Acesso em Maio de 2016.

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, v. 36, n. 4, p. 1165-1188, 2012. Disponível em . Acessado em Junho de 2015.

CURTY, Renata Gonçalves; SERAFIM, Jucenir da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. Revista Informação e Informação. DOI:10.5433/1981-8920. 2016v21n2p307. Londrina, v. 21, n. 2, p. 307–328. Disponível em: <<http://www.uel.br/revistas/informacao/>>. Acesso em jun de 2019.

FERRARI, Luciana Itida. Uma Metodologia para Extração de Informação sobre o Sistema Imunológico. Tese (Doutorado). Coordenação do Programa de pósgraduação em Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro (COPPE/UFRJ), RJ, 2008.

LOTT, Yuri Monnerat and CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. *Perspect. ciênc. inf.* [online]. 2018, vol.23, n.4, pp.117-132. ISSN 1413-9936. Disponível em: <http://dx.doi.org/10.1590/1981-5344/3313>. Acesso em jun de 2019.

PIATETSKY-SHAPIRO, G. FAYYAD, U. An Introduction to SIGKDD and A Reflection on the Term 'Data Mining'. SIGKDD Explorations Volume 13, Issue 2, Page 102, 2011. Disponível em . Acesso em Maio de 2016.

## **Cronograma:**

### **Observação:**

A bibliografia extra será disponibilizada pelo professor no primeiro dia de aula por meio de um link para acesso à biblioteca virtual da disciplina.

### CRONOGRAMA

Início do período letivo: 13/03/2023.

Término do período letivo: 14/07/2023.

Provas finais: 15 a 21/07/2023.

Aulas às quartas-feiras de 18h às 22h

Datas das aulas e conteúdo planejado.

Obs.: podem ocorrer eventuais alterações conforme necessidade, sendo que o registro das aulas dadas, com o conteúdo que foi efetivamente trabalhado, é atualizado e divulgado semanalmente no AVA da disciplina.

Aula 01: 15/03/2023

Aula 02: 22/03/2023

Aula 03: 29/03/2023

Aula 04: 05/04/2023

Aula 05: 12/04/2023

Aula 06: 19/04/2023

Aula 07: 26/04/2023

Aula 08: 03/05/2023

Aula 09: 10/05/2023

Aula 10: 17/05/2023

Aula 11: 24/05/2023

Aula 12: 31/05/2023

Aula 13: 07/05/2023

Aula 14: 14/05/2023

Aula 15: 21/05/2023

Aulas de reposição (se necessário):

28/05/2023, 05 e 12/06/2023

Prova final: 21/07/2023

-----  
Aula 01

Apresentação da disciplina.

Unidade 1: Dados.

Prática: Busca de uma base para o PBL em sites de Dados Abertos.

Aula 02

Unidade 1: (cont.) Dados.

Prática: Software OpenRefine

Aula 03  
Unidade 1: (cont.) Dados.  
Prática: Software OpenRefine

Aula 04  
Unidade 2: Introdução à Ciência de Dados.  
Prática: Software OpenRefine

Aula 05  
Unidade 2: (cont.) Introdução à Ciência de Dados  
Prática: Software OpenRefine

Aula 06  
Unidade 3: Introdução a Banco de Dados  
Prática: Software MySQL Workbench.

Aula 07  
Unidade 3: (cont.) Introdução a Banco de Dados  
Prática: Software MySQL Workbench.

Aula 08  
Exercícios.  
Prática: Software Power BI Desktop.

Aula 09  
Unidade 4: Introdução à Ciência das Redes  
Prática: Software Power BI Desktop.

Aula 10:  
Unidade 4: (cont.) Introdução à Ciência das Redes  
Prática: Software Gephi.

Aula 11:  
Unidade 5: (cont.) Análise de Redes de Informação  
Prática: Software Gephi.

Aula 12:  
Unidade 5: (cont.) Análise de Redes de Informação  
Prática: Software Gephi.

Aula 13:  
Unidade 5: (cont.) Análise de Redes de Informação  
Prática: Software Gephi.

Aula 14:  
Finalização do desenvolvimento do PBL.

Aula 15:  
Apresentações finais do PBL.