



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: 2022 / 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:

Conteúdo Programático:

Unidade 1 - Arquivista de Dados

1.1 Gestão e curadoria de dados

1.2 Competências em literacia de dados

Unidade 2 - Dados e Variáveis

2.1 Dados relacionados com informação e conhecimento

2.2 Categoria de dado quanto a natureza

2.3 Categoria de dado quanto a organização

2.4 Categoria de dado quanto a concepção

2.5 Classificação de variáveis

2.6 Tipos de variáveis quanto ao processo de análise de dados

Unidade 3 - Dados na Internet

3.1 Dados Abertos (Open Data)

3.2 Internet of Events e Big Data

3.3 Princípios FAIR

Unidade 4 - Web de Dados

4.1 Contexto

4.2 Dados Ligados

4.3 Classificação 5 estrelas de Berners-Lee

Unidade 5 - Ciência de Dados (Data Science)

- 5.1 Definição
- 5.2 Cientista de Dados
- 5.3 Exemplos de aplicações de Ciência de Dados
- 5.4 Correlação entre variáveis
- 5.5 Correlação não implica em causalidade
- 5.6 Manipulações tendenciosas de dados
- 5.7 Mineração de Dados (Data Mining)
- 5.8 Data Warehouse
- 5.9 Descoberta de Conhecimento em Bases de Dados (KDD)
- 5.10 Visualização
- 5.11 Metodologias e técnicas da Ciência de Dados

Unidade 6 - Ciência das Redes (Network Science)

- 6.1 Redes Complexas
- 6.2 Redes Sociais
- 6.3 Cronologia, modelos e marcos históricos

Unidade 7 - Análise de Redes Complexas

- 7.1 Algumas aplicações
- 7.2 Fundamentos
- 7.3 Formatação e Inspeção Visual
- 7.4 Conceitos Básicos
- 7.5 Métricas de Centralidade
- 7.6 Coesão, Comunidades e Contágio
- 7.7 Redes Bipartidas (bipartite)
- 7.8 Projeção Bipartida
- 7.9 Software para Análise de Redes Complexas

Unidade 8 - Estudos de Casos

Prática 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.

Desenvolvimento e correção de exercícios e problemas.

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

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.

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 :

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: 18/04/2022
Término do período letivo: 18/08/2022
Provas finais: 19 a 25/08/2022.

Aulas às quartas-feiras de 18h às 22h

Seguem as datas das aulas com o conteúdo planejado, podendo ocorrer eventualmente alterações conforme necessidade.

20/04/2022

Apresentação da disciplina
Unidade 1 - Arquivista de Dados

27/04/2022

Unidade 2 - Dados e Variáveis
Unidade 3 - Dados na Internet
Unidade 4 - Web de Dados

04/05/2022

Unidade 5 - Ciência de Dados (Data Science)
Busca de uma base para o PBL

11/05/2022

Prática com limpeza e preparação de dados
Prática na base de dados escolhida pra o PBL

18/05/2022

Unidade 6 - Ciência das Redes (Network Science)
Prática com redes de referências bibliográficas

25/05/2022

Continuação da prática com redes de referências bibliográficas

01/06/2022

Unidade 7 - Análise de Redes Complexas

08/06/2022

Unidade 8 - Estudos de Casos

Prática na base de dados escolhida pra o PBL

15/06/2022

Unidade 8 - Estudos de Casos

Prática na base de dados escolhida pra o PBL

22/06/2022

Unidade 8 - Estudos de Casos

Prática na base de dados escolhida pra o PBL

29/06/2022

Prática na base de dados escolhida pra o PBL

06/07/2022

Prática na base de dados escolhida pra o PBL

13/07/2022

Prática na base de dados escolhida pra o PBL

20/07/2022

Apresentação do storytelling do PBL completo

27/07/2022

Apresentação do storytelling do PBL completo

03/08/2022

Aula de reposição

10/08/2022

Aula de reposição

17/08/2022

Aula de reposição