



UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC
CENTRO DE CIÊNCIAS DA EDUCAÇÃO – CED
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO – CIN
CAMPUS UNIVERSITÁRIO - TRINDADE
CEP: 88040-970 - FLORIANÓPOLIS - SC
Fone: (048) 3721-4075 E-mail: cin@contato.ufsc.br

PROGRAMA DE ENSINO

1 IDENTIFICAÇÃO

Disciplina: **CIN7915 - Data Science**

Carga Horária: 72 h/a – 4 créditos (2 teóricos e 2 práticos)

Oferta: Optativa para os Cursos de Graduação em Ciência da Informação, Biblioteconomia, Arquivologia e áreas afins

EMENTA

Extração de informação a partir de dados. Definição de domínios de interesse. Segurança de dados. Mineração de dados científicos. Data Warehouse.

2 OBJETIVOS

2.1 Objetivo Geral

Conhecer os fundamentos sobre ciência de dados, assim como a utilização dos dados como fator de vantagem competitiva.

2.2 Objetivos Específicos:

2.2.1 Conhecer os tipos de dados;

2.2.2 Entender os fundamentos de análise exploratória e preditiva;

2.2.3 Estudar os fundamentos da ciência de dados;

2.2.4 Aplicar os principais métodos e ferramentas para análise e ciência de dados.

3 CONTEÚDO PROGRAMÁTICO

3.1 Definição de Projeto

3.1.1 Análise de problemas reais.

3.1.2 Definição de problemas a serem resolvidos com soluções Web.

3.2 Especificação e Modelagem

3.2.1 Levantamento de requisitos do projeto.

3.2.2 Especificação dos requisitos elencados.

3.2.3 Aplicação de Modelagem de bases de dados.

3.2.4 Modelagem de interfaces.

3.2.5 Modelagem de controles Web.

3.3 Implementação

3.3.1 Implementação de bancos de dados com tecnologias emergentes.

3.3.2 Implementação de modelos de controle com tecnologias emergentes.

3.3.2 Implementação de modelos de interface com tecnologias emergentes.

AULA (semana)	DATA	ASSUNTO
1	30/07	Apresentação da disciplina, Introdução à Data Science, Discussão sobre Data Science. Artigo: Cao, Longbing. " Data science: a comprehensive overview. " <i>ACM Computing Surveys (CSUR)</i> 50.3 (2017): 43.
2	06/08	Introdução ao Ambiente de Desenvolvimento, Python, spyder, Anaconda. Pré-requisitos para o desenvolvimento de Data Science com Python.
3	13/08	Introdução à Numpy
4	20/08	Introdução à Pandas
5	27/08	Pré-processamento de Dados
6	03/09	Introdução à visualização de Dados
7	10/09	Dia Não Letivo
8	17/09	Dia não Letivo
9	24/09	Análise de Dados
10	01/10	Análise de Dados
11	08/10	Ler Dados de csv online e de banco de dados. Apresentação de Datasets (UCI e Governo) Definição de projeto.
12	15/10	Dia Não Letivo
13	23/10	Acompanhamento de Projeto
14	30/10	Acompanhamento de Projeto
15	06/11	Acompanhamento de Projeto
16	13/11	Acompanhamento de Projeto
17	20/11	Acompanhamento de Projeto
18	27/11	Acompanhamento de Projeto
19	04/12	Entrega do Projeto e os Resultados Escritos em Artigo.

Obs: O calendário está sujeito a pequenos ajustes de acordo com as necessidades das atividades desenvolvidas.

XII. FERIADOS PREVISTOS PARA O SEMESTRE 2018.2:

DATA	
07/08	Independência do Brasil
14/08	Dia não letivo.
12/10	Nossa Senhora Aparecida

6. METODOLOGIA

Aulas expositivas, com discussão em classe dos tópicos apresentados e exercícios práticos e/ou escritos. Teremos como apoio das disciplinas o **Moodle** institucional da UFSC (<http://moodle.ufsc.br>) e o mesmo será utilizado como instrumento de suporte da disciplina, devendo portanto, ser acessado todas as semanas pelos alunos.

7. AVALIAÇÃO

Conforme a Resolução do Conselho Universitário 017/Cun/97, de 30/09/97:

§ A frequência mínima obrigatória é de 75% das aulas.

§ Em caso de falta em dia de avaliação (prova, apresentação de trabalho ou minisseminário), encaminhar justificativa e pedido formal à Chefia do Departamento de Ciência da Informação, no prazo de três (3) dias úteis.

§ O aluno que obtiver nota final entre 3,0 e 5,5 e frequência suficiente poderá, ao final do semestre, realizar uma prova de recuperação de todo o conteúdo.

§ A nota final será calculada a partir da média simples entre as três provas (se a média for $\geq 5,75$) ou obtida pela média obtida entre a média das três provas e a obtida na prova de recuperação.

Avaliações:

- Será realizado um projeto individual e a nota será dividida em Trabalho Escrito (TE) e duas Apresentações de Projetos (AP). O Trabalho Escrito será equivalente a 70% da nota total. Cada Apresentação será equivalente a 15% da nota.

- A média final (MF) será calculada conforme a seguinte fórmula:

$$\mathbf{MF = (TE * 0,7) + (((AP1 + AP2) / 2) * 0,3)}$$

Se $3 \leq \mathbf{MF} \leq 5 \Rightarrow \mathbf{MF = (((TE * 0,6) + (AP * 0,4)) + REC)/2}$, onde REC = prova de recuperação

8 BIBLIOGRAFIA

8.1 Bibliografia básica

AMARAL, Fernando. Introdução à Ciência de Dados: Mineração de Dados e Big Data. Rio de Janeiro: Alta Books, 2016. 320p.

DA SILVA, Leandro; PERES, Sarajane; BOSCARIOLI, Clodis. Introdução à Mineração de Dados com Aplicações em R. Rio de Janeiro: Elsevier, 2016. 277p.

DE CASTRO, Leandro; FERRARI, Daniel. Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações. São Paulo: Saraiva, 2016. 351p.

PROVOST, Foster; FAWCETT, Tom. Data science for business. Sebastopol: O'Reilly, c2013. xxi, 386 p. ISBN 9781449361327.

8.1 Bibliografia Complementar

ATZENI, Paolo. Database systems: concepts, languages & architectures. London: McGraw Hill, c2000. xxiii, 612p. ISBN 007235872.

DATE, C. J. Introdução a sistemas de bancos de dados. Rio de Janeiro: Campus, 2004. 865p. ISBN 8535212736.

FINLAY, Steven. Predictive analytics, data mining, and big data: myths, misconceptions and methods. New York: Palgrave Macmillan, c2014. xii, 248 p. ISBN 9781137379276.

HURWITZ J. et al. Big Data for Dummies. New Jersey (USA): John Wiley & Sons, Inc., 2013. ISBN 978-1-118-50422-2.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. Big data: a revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Company, c2013. 252 p. ISBN 9780544227750.

MILLER, Harvey J.; HAN, Jiawei. Geographic data mining and knowledge discovery. 2nd ed. Boca Raton: CRC Press, c2009 458 p. ISBN 9781420073973.

NISBET, Robert; ELDER, John; MINER, Gary. Handbook of statistical analysis and data mining applications. Amsterdam: Elsevier; London: Academic Press, 2009. xxxiv, 824 p. ISBN 9780123747655.