

**UNIVERSIDADE DE SÃO PAULO - ICMC**

Departamento de Ciências de Computação

SCC-0275 – Introdução à ciência de dados – Graduação - 2º Sem /2018

PROFa: Roseli Aparecida Francelin Romero

Estagiário PAE: Daniel Moreira Cestari

Aula	Dia	Conteúdo
1	02/08	Apresentação: a disciplina Introdução à ciências de dados, Objetivos e Avaliação. Formação e onde trabalha cientista de dados. Big Data. Ciência de Dados. Introdução ao python.
2	09/08	Knowledge discovery in databases (KDD): Metodologias, KDD, MD. Descoberta de Conhecimento em Bases de Dados. Etapas de KDD. Mineração de Dados. Aplicações. Descrição de dados. Qualidade de dados. Apresentação das ferramentas e bibliotecas.
	16/08	Não Haverá Aula – Semana da Computação
3	23/08	Exploração de dados – Parte I: Dados. Caracterização de dados (Objetos e atributos; Tipos de dados). Exploração de dados (Dados univariados; Dados multivariados; Visualização). ATENÇÃO: Os alunos de Graduação: entregar o Projeto I
4	30/08	Exploração de dados – Parte II: Medidas de localização (boxplot e boxplot modificado). Medidas de dispersão (intervalo, variância). Medidas de distribuição (momentos estatísticos, histograma). Dados multivariados (frequência, localização, dispersão e distribuição). ATENÇÃO: Os alunos de Graduação: entregar o Projeto II
	06/09	Não Haverá Aula – Semana da Pátria
5	13/09	Exploração de dados – Parte III: Outras formas de sumarizar os dados. Visualização gráfica (histograma, diagrama de torta, scatter plot, faces de Chernoff). Infográficos. ATENÇÃO: Os alunos de Graduação: entregar o Projeto III
6	20/09	Pré-Processamento de dados – Parte I: Pré-processamento. Qualidade dos dados. Limpeza dos dados (valores ausentes, descarte, imputação). Valores inconsistentes. Objetos redundantes. Ruído. Outliers. ATENÇÃO: Os alunos de Graduação: entregar o Projeto VI
7	27/09	Pré-Processamento de dados – Parte II: Dados desbalanceados (balanceamento artificial, modelos para classes). Transformação dos dados. Qualitativos para quantitativos. Quantitativos para qualitativos. Transformações de atributos (funções simples, normalização, padronização). Conversão de valores numéricos. Tradução. ATENÇÃO: Os alunos de Graduação: entregar Projeto V
8	04/10	Prova 1
9	11/10	Experimentos – Parte I: Modelagem. Desempenho (preditivo, custo). Amostragem (hold-out, reamostragem). Classificação binária. Medidas de avaliação. ATENÇÃO: Os alunos de Graduação: entregar Projeto VI
10	18/10	Experimentos – Parte II: Acurácia. Precisão. Revocação. Sensibilidade. Especificidade. Medida-F. Outras medidas. Gráficos ROC (curva ROC, AUC). Teste de hipótese. Teste hierárquico bayesiano. ATENÇÃO: Os alunos de Graduação: entregar Projeto VI
11	25/10	Exercícios
12	01/11	Critérios de distância: Aprendizado baseado em instâncias. 1-vizinho mais próximo. Medidas de distância. Similaridade e dissimilaridade. Proximidade. K-vizinhos mais próximos.
13	08/11	Exercícios
	15/11	Não haverá Aula - Proclamação da República
14	22/11	Prova 2

CRITERIO DE AVALIAÇÃO:

$MF = 0.5 * MP + 0.5 * MT$, ONDE $MP = 0.4 * P1 + 0.6 * P2$ E $MT = \text{MÉDIA ARITMÉTICA DOS TRABALHOS}$, SENDO QUE MP E $MT \geq 5.0$ PARA A APROVAÇÃO. Do contrário, será considerado com $MF = \min\{MP, MT\}$

BIBLIOGRAFIA:

- LIVRO. David J. Hand, Heikki Mannila, Padhraic Smyth. (2001). Principles of Data Mining. Capítulos 1, 2, 3, 4, 5.
- LIVRO. Tan, P.-N.; Steinbach, M.; Kumar, T. (2005). Introduction to Data Mining.
- LIVRO. Paolo Giudici, Silvia Fugini. (2009). Applied Data Mining for Business and Industry. Capítulos 1, 2, 3, 6.
- LIVRO. Philipp Janert. (2010). Data Analysis with Open Source Tools.
- LIVRO. Faceli, Katti; Lorena, Ana Carolina; Gama, João ; de Carvalho, A. C. P. L. F. (2011). Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina, 1. ed.
- LIVRO. Han, J.; Kamber, M.; Pei, J. (2011). Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems). Capítulos 1, 2, 3, 4, 5.
- LIVRO. Ian Witten, Eibe Frank, Mark Hall. (2011). Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition.
- LIVRO. (2011). Statistical and Machine-Learning Data Mining. Capítulos 1, 2.
- LIVRO. Provost, F.; Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking.
- LIVRO. Cathy O'Neil, Rachel Schutt. (2013). Doing Data Science. Capítulos 1, 2, 7, 9.
- LIVRO. Pedro A. Morettin, Wilton de O. Bussab. (2014). Estatística básica.

EBOOK. Charu C. Aggarwal. (2015). Data Mining: The Textbook. Capítulos 1, 2.

EBOOK. Thomas A. Runkler. (2016). Data Analytics. Capítulos 1, 2, 3, 4.

EBOOK. Doug Rose. (2016). Data Science. Capítulos todos exceto da Parte II.

EBOOK. Rajendra Akerkar, Priti Srinivas Sajja. (2016). Intelligent Techniques for Data Science. Capítulos 1, 2.

EBOOK. Max Bramer. (2016). Principles for Data Mining. Capítulos 1, 2, 7, 8.

EBOOK. Steven S. Skiena. (2017). The Data Science Design Manual. Capítulo 1.

BIBLIOGRAFIA COMPLEMENTAR:

LIVRO. Tom Mitchell. (1997). Machine Learning. Capítulos 1 e 2.

EBOOK. Huan Liu, Hiroshi Motoda. (1998). Features Selection for Knowledge Discovery. Capítulos 1, 2, 6, 7.

EBOOK. Mohamed Medhat Gaber. (2012). Journeys to Data Mining.

EBOOK. Verónica Bolón-Canedo, Noelia Sánchez-Marroño, Amparo Alonso-Betanzos. (2015). Feature Selection For High-Dimensional Data. Capítulos 1, 2, 5.

EBOOK. Wilfried Grossmann, Stefanie Rinderle-Ma. (2015). Fundamentals of Business Intelligent. Capítulos 1, 2, 4.

EBOOK. Brian Steele, John Chandler, Swarna Reddy. (2016). Algorithms for Data Science. Prefácio, capítulos 1, 2, 3, 5, 7.

EBOOK. Sherif Sakr. (2016). Big Data 2.0 Processing Systems. Capítulo 1.

EBOOK. José María Cavanillas, Edward Curry, Wolfgang Wahlster. (2016). New Horizons For a Data-Driven Economy. Capítulos 1, 3, Part II.

MOOC. Yaser S. Abu-Mostafa. Learning From Data. (<https://work.caltech.edu/telecourse.html>)

MOOC. Nas plataformas edX (<https://www.edx.org/>) e Coursera (<https://www.coursera.org/>) procurar por data science.

EBOOK. Fabrizio Romano. (2015). Learning Python. (<https://www.packtpub.com/packt/free-ebook/learning-python>)

WEBSITE. The Python Tutorial. (<https://docs.python.org/3/tutorial/>)

WEBSITE. 15 Free eBooks to Learn Python. (<https://codeburst.io/15-free-ebooks-to-learn-python-c299943f9f2c>)