



Plano de Disciplina

Identificação				
CURSO Mestrado Profissional em Tecnologia da Informação			CAMPUS João Pessoa	
DISCIPLINA Mineração de Dados			CÓDIGO DA DISCIPLINA 30630	
PRÉ-REQUISITO Conhecimentos sobre gerenciamento de dados e programação.				
UNIDADE CURRICULAR (OBRIGATÓRIA, OPTATIVA, ELETIVA) Eletiva			SEMESTRE 2021.2	
DOCENTES RESPONSÁVEIS Damires Yluska de Souza Fernandes				
Carga Horária				
TEÓRICA 24	PRÁTICA 24	EaD ¹ 12	CARGA HORÁRIA SEMANAL: 4	CARGA HORÁRIA TOTAL 60

Ementa

Introdução à Mineração de Dados e Ciência de Dados; Cenários de aplicação; Processos baseados no KDD e CRISP-DM; Aprendizado supervisionado e não-supervisionado; Tarefas de Mineração de Dados com foco em Classificação e Clusterização; Avaliação, análise e interpretação de resultados. Projeto aplicado ao tema de mestrado do estudante.

Objetivos

Estudar, discutir e aplicar conceitos e técnicas de mineração de dados e dos processos de KDD e CRISP-DM com vistas à identificação de padrões importantes e não óbvios em conjuntos de dados normalmente volumosos. A motivação para a pesquisa e estudos nesta área é o desafio em lidar eficientemente com conjuntos de dados em diversos domínios de aplicação de modo a extrair padrões e análises que possam agregar valor e ajudar em tomadas de decisão.

Objetivos específicos:

- Caracterizar e entender a Ciência de Dados, a Mineração de Dados e conceitos afins;
- Compreender e aplicar os principais passos de processos de descoberta de conhecimento em conjuntos de dados;
- Compreender as principais tarefas de mineração de dados;
- Conhecer e aplicar as principais técnicas de mineração de dados baseadas em aprendizado supervisionado no contexto de classificação;
- Conhecer as principais técnicas de mineração de dados baseadas em aprendizado não supervisionado no contexto de clusterização;
- Conhecer e aplicar técnicas de avaliação de resultados de modelos de aprendizado gerados.

¹ Para a oferta de disciplinas na modalidade à distância, integral ou parcial, desde que não ultrapasse 20% (vinte por cento) da carga horária total do curso, observar o cumprimento da Portaria MEC nº 1.134, de 10 de outubro de 2016.



Conteúdo Programático – Presencial/Remoto

Tópico	Carga horária
Introdução à Ciência de Dados, Mineração de Dados e conceitos afins	4
Mineração de dados, KDD e CRISP-DM (etapas), tipos de dados, de aprendizado de máquina e análises.	4
Pre-processamento de dados: limpeza, integração, transformação, redução e discretização de dados	8
Tarefas de mineração de dados e aprendizado supervisionado	12
Tarefas de mineração de dados e aprendizado não supervisionado	8
Técnicas de avaliação	4
Desenvolvimento de projeto e de artigo: seleção e preparação de conjunto de dados, estudo e aplicação de método(s) de classificação, avaliação e análise dos resultados	8

Conteúdo Programático - EAD

Tópico	Carga horária
Desenvolvimento de projeto	8
Escrita de artigo	4

Total	60
--------------	-----------

Metodologia de Ensino

Presencial/remota

- Aulas e atividades síncronas e assíncronas expositivas/discursivas/dialogadas/executórias
- Leituras, estudos e atividades complementares (artigos, trabalhos, documentações, uso de ferramentas, etc.)
- Participações em atividades: exercícios, fóruns, pesquisas propostas, discussões/apresentações de trabalhos/artigos aplicados

EAD

- Planejamento e acompanhamento de projeto da disciplina. Acompanhamento para cada discente/grupo e entrega de milestones em datas pré estabelecidas.
- Planejamento e acompanhamento de escrita de artigo a respeito dos resultados do projeto desenvolvido. Entregas de etapas do artigo segundo deadlines pré-definidos.

Recursos Didáticos

Ambiente virtual de aprendizagem institucional, ferramentas, periódicos, artigos, acesso à Internet, materiais de aula, videoaulas, webaulas, atividades remotas.

Critérios de Avaliação

A avaliação dar-se-á com base em três marcos: (i) Atividades contínuas (pesquisas e exercícios) ao longo da disciplina (**T1**); (ii) Projeto prático desenvolvido (**P1**); (iii) Artigo escrito e submetido resultante da pesquisa/projeto realizado (**A1**). Todas as atividades terão sub marcos e deadlines definidos a serem cumpridos. Ao final, a nota da disciplina será **Nota = (T1*3 + P1*4 + A1*3)/10**.



Bibliografia

1. HAN, J.; KAMBER, M.; PEI, J. Data Mining Concepts and Techniques. 3. ed. Waltham (USA): Morgan Kaufmann, 2012.
2. DE CASTRO, Leandro Nunes; FERRARI, Daniel Gomes. Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações. Ed. Saraiva. 2016.
3. TAN P., STEINBACK M. e KUMAR V. Introduction to Data Mining, Pearson, 2013.
4. GOLDSCHMIDT R., PASSOS E. Datamining - Conceitos, Técnicas, Algoritmos, Orientações e Aplicações. Editora Campus, 2015.
5. ZAKI M., MEIRA, JR. W., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Edition, Cambridge University Press, March 2020. ISBN: 978-1108473989.
6. AMARAL, F. Introdução à Ciência de Dados. Editora Alta Books, 2016.
7. ALPAYDIN, Elthem. Introduction to Machine Learning 2nd Edition. MIT Press, Massachusetts, USA, 2 edition, 2010.
8. MITCHELL, Thom. Machine Learning. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
9. WITTEN Ian, FRANK Eibe. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.
10. KELLEHER, John D.; MAC NAMEE, Brian; D'ARCY, Aoife. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge: MIT Press, 2015.
11. ELMASRI, R.; NAVATHE S. B. Sistemas de Banco de Dados. 6 ed. São Paulo: Pearson Education, 2011.
12. DONG, X. L.; SRIVASTAVA, D. Big Data Integration. [S.l.]: Morgan & Claypool. 2015.
13. Artigos relacionados aos tópicos de interesse na disciplina, selecionados a partir de publicações em revistas e eventos de alto impacto.
14. Documentações diversas como <https://scikit-learn.org/stable/> e https://pandas.pydata.org/docs/user_guide/index.html

Observações

(Nenhuma)