



## 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 2023.2	
DOCENTES RESPONSÁVEIS Damires Souza e Thiago Moura				
Carga Horária				
TEÓRICA 24	PRÁTICA 24	EaD <sup>1</sup> 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; Algoritmos KNN, Árvore de Decisão, Modelos Lineares, Redes Neurais Artificiais, Ensembles, K-Means; Seleção de features; Balanceamento de dados; Avaliação, análise e interpretação de resultados. Projeto aplicado ao tema de mestrado/interesse do estudante.

### Objetivos

Estudar, discutir e aplicar conceitos e técnicas de mineração de dados e de aprendizado de máquina dentro de processos como o 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 Ciência de Dados, Mineração de Dados, Aprendizado de Máquina e conceitos afins;
- Compreender e aplicar os principais passos de processos de descoberta de conhecimento em conjuntos de dados;
- Compreender e aplicar os principais algoritmos de mineração de dados e de aprendizado de máquina;
- Conhecer e aplicar técnicas de avaliação de resultados de modelos de aprendizado gerados.
- Desenvolver projeto aplicado a um problema/cenário real.

<sup>1</sup> 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 a Ciência de Dados, Mineração de Dados, Aprendizado de Máquina e conceitos afins	4
Mineração de dados, KDD e CRISP-DM (etapas), tipos de dados, de aprendizado de máquina e análises.	2
Preprocessamento de dados: limpeza, transformação, redução, seleção, balanceamento e discretização de dados	12
Algoritmos de aprendizado supervisionado	18
Algoritmos de aprendizado não supervisionado	4
Técnicas de separação e avaliação	4
Desenvolvimento de projeto aplicado: seleção e preparação de conjunto de dados, estudo e aplicação de método(s) de AM, avaliação e análise dos resultados	4

### Conteúdo Programático - EAD

Tópico	Carga horária
Desenvolvimento de projeto	12

<b>Total</b>	<b>60</b>
--------------	-----------

### Metodologia de Ensino

#### Presencial

- Aulas e atividades expositivas/discursivas/dialogadas/executórias
- Leituras, estudos e atividades complementares (artigos, trabalhos, documentações, implementações, 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.

### Recursos Didáticos

Sala de aula, ambiente virtual de aprendizagem institucional, ferramentas, periódicos, artigos, acesso à Internet, materiais de aula, vídeos.

### Critérios de Avaliação

A avaliação dar-se-á com base em dois marcos: (i) Atividades contínuas (pesquisas e exercícios) ao longo da disciplina (**A1**); e (ii) Projeto prático desenvolvido e apresentado (**P1**). Todas as atividades terão sub marcos e deadlines definidos a serem cumpridos. Ao final, a nota da disciplina será a média aritmética entre **A1** e **P1**.

### 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. Artigos relacionados aos tópicos de interesse na disciplina, selecionados a partir de publicações em revistas e eventos de alto impacto.
13. Documentações diversas como <https://scikit-learn.org/stable/> e [https://pandas.pydata.org/docs/user\\_guide/index.html](https://pandas.pydata.org/docs/user_guide/index.html)

## Observações

(Nenhuma)