

Conteúdo Programático, Bibliografia (indicação opcional) e Sistematização da Prova Prática (quando houver)

Edital UFRJ nº 875, de 19 de setembro de 2025

Haverá Prova Prática: () Sim (X) Não

Unidade			
Código da Opção de Vaga	MC-002	Departamento ou Programa / Setorização Definitiva	Programa de Engenharia de Sistemas e Computação / Ciência de Dados
Conteúdo Programático	<ol style="list-style-type: none">1. Fundamentos de Estatística e Probabilidade: distribuições de probabilidade, momentos, estimação, testes de hipóteses, regressão linear e métodos de Monte Carlo.2. Fundamentos de Aprendizado de Máquina: paradigmas de aprendizado, classificação, regressão, agrupamento, avaliação de Modelos e técnicas de ensemble.3. Modelos Avançados de Aprendizado: redes neurais convolucionais, recorrentes, e transformers, técnicas de redução de dimensionalidade, aprendizado semi-supervisionado e auto-supervisionado.4. Mineração e Exploração de Dados: pré-processamento, integração, descoberta de conhecimento, seleção de atributos, séries temporais, análise de textos e dados multimodais.5. Busca e Recuperação de Informação: modelos clássicos e probabilísticos, indexação, feedback de relevância, recuperação multimodal e métricas de avaliação.6. Processamento de Linguagem Natural: fundamentos, representação de texto, modelos probabilísticos, embeddings, transformers e aplicações.7. Engenharia de Software para Ciência de Dados: boas práticas de programação, testes, versionamento, containers, CI/CD, MLOps, experimentação, implantação e monitoramento de modelos.8. Processamento em Larga Escala: computação paralela e distribuída de dados, sistemas de streaming, computação em nuvem aplicada à ciência de dados e programação científica.9. Big Data: conceitos fundamentais, arquiteturas, processamento em lote e fluxo, escalabilidade, tolerância a falhas, governança e segurança.10. Engenharia de Dados: Modelagem Conceitual, fusão e integração de dados, consultas a dados heterogêneos.11. Processos e Modelos de Gestão do Conhecimento: fundamentos, ciclo de vida, modelos clássicos, tecnologias de apoio, métricas e indicadores, e integração com ciência de dados.12. Ética e Governança em Ciência de Dados: vieses, reprodutibilidade, explicabilidade de modelos, responsabilidade no uso de dados, e conformidade com GDPR e LGPD.		
Bibliografia (indicação opcional)	<ol style="list-style-type: none">1. HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. Springer, 2009.2. MURPHY, K. P. Probabilistic Machine Learning: An Introduction. MIT Press, 2022.3. GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. MIT Press, 2016.4. BAEZA-YATES, R.; RIBEIRO-NETO, B. Modern Information Retrieval: The Concepts and Technology behind Search. Addison-Wesley, 2011.5. DALKIR, K. Knowledge Management in Theory and Practice, 3rd edition. MIT Press, 2017.		

	<p>6. OSZU, T.; VALDURIEZ, P. Principles of Distributed Database Systems, 4th edition, Springer, 2020.</p> <p>7. KLEPPMANN, M. Designing Data-Intensive Applications. O'Reilly, 2017.</p> <p>8. JURAFSKY, D.; MARTIN, J. H. Speech and Language Processing, 3rd edition, 2025 (online).</p> <p>9. FELDERER, M.; TRAVASSOS, G.H. Contemporary Empirical Methods in Software Engineering, Springer, 2020.</p> <p>10. BIRD, C.; MENZIES, T.; ZIMMERMANN, T. The Art and Science of Analyzing Software Data, Elsevier, 2015.</p> <p>11. WÖHLIN, C.; RUNESON, P.; HÖST, M. et al. Experimentation in Software Engineering. 2nd edition, 2024.</p>
Sistematização da Prova Prática	