

# **2º Encontro da SBPC em MS/ XI ENEPEX / XIX ENEPE/ 22ª SNCT - UEMS / UFGD 2025**

## **INVESTIGANDO TÉCNICAS DE TESTE PARA PROJETOS DE CIÊNCIA DE DADOS**

**Instituição:** Universidade Estadual de Mato Grosso do Sul

**Área temática:** Ciência da Computação

**RECALCATTI, João Pedro<sup>1</sup>** ([rgm49117@comp.uems.br](mailto:rgm49117@comp.uems.br)); **PRATES, Jorge Marques<sup>2</sup>** ([jprates@uems.br](mailto:jprates@uems.br)).

<sup>1</sup> – Discente do Curso de Sistemas de Informação, UEMS;

<sup>2</sup> – Docente do Curso de Sistemas de Informação, UEMS.

A crescente popularidade da Ciência de Dados na indústria de software tem impulsionado o desenvolvimento de soluções complexas que lidam com grandes volumes de dados. Contudo, essa rápida expansão não foi acompanhada pela adoção sistemática de práticas rigorosas de Engenharia de Software, especialmente no que tange aos testes. Essa lacuna resulta em sistemas baseados em dados com problemas de confiabilidade, manutenibilidade e robustez, comprometendo a qualidade e a eficácia das soluções. A aplicação de técnicas de teste consolidadas é, portanto, fundamental para garantir a integridade e a precisão das análises e modelos desenvolvidos. Este projeto teve como objetivo geral investigar técnicas de teste de software para avaliar sua viabilidade e relevância quando aplicadas no contexto de projetos de Ciência de Dados. Como objetivos específicos, buscou-se compreender as necessidades particulares de teste dessa área emergente e analisar as práticas de teste de software existentes para identificar aquelas que poderiam ser adaptadas para melhorar a qualidade, a confiabilidade e a eficácia das soluções baseadas em dados, preenchendo uma importante lacuna de pesquisa. A abordagem metodológica adotada foi multifásica, iniciando com uma revisão bibliográfica aprofundada sobre Teste de Software e Ciência de Dados. Seguiu-se um estudo da linguagem de programação *Python* e da biblioteca Pandas para adquirir as habilidades necessárias para o desenvolvimento de projetos na área. Posteriormente, realizou-se um levantamento e análise sistemática das técnicas de teste aplicáveis a projetos de Ciência de Dados, culminando em um estudo de viabilidade que envolveu a implementação de casos de teste em um projeto prático para identificar melhores práticas e lições aprendidas. Resultados; A principal contribuição deste trabalho foi o desenvolvimento de um "Catálogo de Técnicas de Teste para Projetos de Ciência de Dados". Este catálogo sistematiza diversas abordagens, desde testes de caminho correto e incorreto até o uso de técnicas avançadas como *mocks*, *fixtures*, parametrização e validação de esquemas de dados. A eficácia e aplicabilidade do catálogo foram demonstradas por meio da aplicação das técnicas em funções de um estudo de caso prático, que utiliza dados públicos de saúde. Nesse contexto, as técnicas foram aplicadas com sucesso para validar funções de pré-processamento e filtragem de dados, garantindo a integridade e a correção de um pipeline de análise. O estudo concluiu que a integração de princípios de teste de software em fluxos de trabalho de Ciência de Dados é não apenas viável, mas essencial para aumentar a robustez e a confiabilidade dos sistemas. O catálogo desenvolvido serve como um guia prático para pesquisadores e desenvolvedores, oferecendo um *framework* estruturado para a validação de projetos de dados.

**PALAVRAS-CHAVE:** Engenharia de Software, Teste, Ciência de Dados.

**AGRADECIMENTOS:** Agradecemos à Diretoria de Relações Internacionais e à Universidade Estadual de Mato Grosso do Sul (UEMS) pelo apoio institucional e financeiro. Agradecemos também ao professor Dr. Gautam Srivastava pela colaboração durante a mobilidade internacional, que foi essencial para o aprofundamento da pesquisa.