

# AUTOMAÇÃO DO PROCESSO DE COLETA E CLASSIFICAÇÃO DAS NOTÍCIAS PUBLICADAS NO SITE DO IF GOIANO COM ÊNFASE NA QUANTIDADE DE ACESSOS

**Rafael Quintino Fonseca Guimarães<sup>1</sup>; Jesmmer da Silveira Alves<sup>2</sup>;**

<sup>1</sup>Ciência da Computação, Instituto Federal Goiano – Campus Morrinhos, [rafael.fonseca@estudante.ifgoiano.edu.br](mailto:rafael.fonseca@estudante.ifgoiano.edu.br);

<sup>2</sup>Ciência da Computação, Instituto Federal Goiano – Campus Morrinhos, [jesmmer.alves@ifgoiano.edu.br](mailto:jesmmer.alves@ifgoiano.edu.br).

**RESUMO:** Este estudo tem como propósito coletar e classificar de forma automática as publicações no site do Instituto Federal Goiano em categorias como extensão, comunicados, premiações, campanhas, editais, literatura e avaliações, identificando as mais acessadas e suas razões de sucesso. Utilizamos técnicas de Web Scraping em Python com Selenium para extrair dados do site. A automação foi otimizada com a plataforma Airflow, permitindo agendamento diário sem intervenção manual. Para inferir temas, aplicamos a inteligência artificial ChatGPT, analisando títulos e descrições das publicações, conforme as categorias mencionadas. A avaliação da quantidade de acessos foi realizada por meio da linguagem Python e biblioteca Pandas, fornecendo insights valiosos para melhor compreensão dos padrões de interesse e preferências dos usuários. Técnicas de análise e visualização de dados foram usadas para extrair informações importantes sobre as publicações.

**Palavras-chave:** análise de dados; automação com airflow; python; selenium; web scraping.