

# ESTUDOS APLICADOS À DATA SCIENCE COM FOCO EM DADOS PÚBLICOS

Leticia da Silva Rosa<sup>1</sup> (UNIP, Bolsista PIBIC/CNPq)  
Rafael Duarte Coelho dos Santos (CTE/LAC/INPE, Orientador)

## RESUMO

Este trabalho, iniciado em agosto de 2018, tem como objetivo o aprendizado e aprofundamento em técnicas utilizadas na área de Ciência de Dados, para sua aplicação em projetos que possuam a necessidade de obter informações contidas em espaços não tão claros para sua captura. Dona de um crescimento notório nos últimos anos, a área de Data Science vem ganhando espaço devido a grande quantidade de dados existentes, dos quais podem ser extraídas informações para diversas frentes. Uma delas é a dos Dados Públicos, que ganhou visibilidade a partir da Lei de Acesso à Informação, em vigor desde 2012. Ela regulamenta que informações públicas dos órgãos e entidades do governo, em todas suas esferas, sejam recebidas por pessoas físicas ou jurídicas. Com base nos fatos apontados, um projeto foi desenvolvido utilizando a plataforma do DATASUS, onde dados de mortalidade dos desastres climáticos mais importantes do Brasil foram extraídos e analisados. O trabalho mostra as dificuldades enfrentadas desde a obtenção até a manipulação dos dados, de que maneira que eles foram disponibilizados, qual foi o impacto gerado no produto e o que pode ser extraído dessa experiência. O projeto conta com a participação de um trabalho de doutorado que deu origem a temática dos dados extraídos, da doutoranda Gabriela Couto, cujo foco é analisar os dez maiores desastres climáticos ocorridos na história brasileira através das lentes do gênero. Isso possibilitou o desenvolvimento de um olhar mais apurado para as situações evidenciadas e os respectivos resultados apresentados. Ferramentas como as linguagens Python, R e SQL, notebooks, e conceitos de Web Scraping e Data Mining foram utilizadas para o estudo. Devido às variadas formas que o dado foi encontrado, é possível notar os desafios apresentados para os profissional da área, que não é capaz de prever qual será o cenário encontrado para a extração das informações necessárias. Espera-se dar continuidade ao projeto de forma a proporcionar maior embasamento na área, possibilitando que novas vertentes venham a ser trabalhadas e que processos de análise, automatização e extração dos dados sejam implantados e melhorados.

<sup>1</sup> Aluna do Curso de Ciência da Computação - **E-mail:** [leticiadasilva.contato@gmail.com](mailto:leticiadasilva.contato@gmail.com)

<sup>2</sup> Prof<sup>o</sup> Doutor da Divisão do Laboratório de Computação e Matemática Aplicada - **E-mail:**

**rafael.santos@inpe.br**