



<b>Evento</b>	Salão UFRGS 2024: SIC - XXXVI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2024
<b>Local</b>	Virtual
<b>Título</b>	Sistematização do uso de web scraping e API's na coleta de dados secundários para estudos sobre sustentabilidade empresarial
<b>Autor</b>	HÉRON HERMANN OLIVEIRA
<b>Orientador</b>	ALEJANDRO GERMAN FRANK

## **Sistematização do uso de web scraping e API's na coleta de dados secundários para estudos sobre sustentabilidade empresarial**

As pesquisas sobre o impacto de fatores ambientais na sustentabilidade empresarial apontam uma necessidade de sistematizar métodos para coletar dados que permitam analisar a correlação entre variáveis relevantes, como 'fatores ambientais' e 'sustentabilidade empresarial' a partir do uso de dados secundários. A pesquisa com dados secundários é promissora para analisar tais impactos com uso de ferramentas estatísticas mais robustas que permitem reduzir a influência de fatores endógenos e os vieses de resposta tipicamente presentes em fontes primárias. Porém, há uma necessidade de maior clareza quanto aos procedimentos que facilitem a coleta e integração desses tipos de dados secundários no âmbito da sustentabilidade. Este trabalho teve com objetivo métodos de coleta para dados secundários a partir do uso de tecnologias de automação de dados, como web scraping, APIs e Machine Learning para mineração de dados, visando a extração e transferência de informações úteis para pesquisas acadêmicas no tema. Essas técnicas foram utilizadas, ou testadas, por meio de linguagens de programação e softwares. Além disso, foram investigadas e documentadas maneiras adequadas de organizar e fornecer esses dados aos interessados em usá-los como recurso de pesquisa. Como resultado, o trabalho sistematiza formas de usos, de web scraping e APIs. Foi desenvolvida uma simulação da aplicação dessas técnicas em amostras de dados, e obtido as principais variáveis para a modelagem de fatores ambientais na sustentabilidade. Avanços na pesquisa sobre Machine Learning também foram registrados, sendo que ainda necessitam de análises adicionais para sua sistematização como método de coleta de dados.

Héron Hermann Oliveira<sup>1</sup>

Alejandro Germán Frank<sup>2</sup>

<sup>1</sup> Bolsista PIBIC CNPq, Universidade Federal do Rio Grande do Sul,

E-mail:heron.hhoh@gmail.com

<sup>2</sup> Professor Orientador, Núcleo de Engenharia Organizacional (NEO), Universidade Federal do Rio Grande do Sul