

Inteligência Artificial Generativa na Análise de Relatórios Ambientais: Automatização e Predição de Emissões Atmosféricas na Região Sul Brasileira

Bruno dos Santos Guterres¹; Andrea Cristina Marin²; Marlon Cavalcante Maynard³

¹ Centro Universitário Senac (bruno.sguterres@senacsp.edu.br)

² Centro Universitário Senac (andrea.cmarin@sp.senac.br)

³ Centro Universitário Senac (marlon.cmaynard@sp.senac.br)

Resumo: A intensificação das atividades industriais na região Sul do Brasil tem ampliado as emissões de poluentes atmosféricos, demandando sistemas de monitoramento mais ágeis para proteger a saúde pública e garantir metas de sustentabilidade. Entretanto, o monitoramento da qualidade do ar no Brasil é frequentemente dificultado pela complexidade de relatórios técnicos anuais, cujos dados estão dispersos em documentos não estruturados, agravado pela indisponibilidade pública de dados em alguns estados, limitando a agilidade de órgãos fiscalizadores na tomada de decisões. Este trabalho apresenta uma metodologia que utiliza Inteligência Artificial (IA) Generativa como elemento central de um processo integrado e automatizado para análise de dados ambientais. O objetivo é demonstrar como modelos de linguagem de grande escala (Large Language Models — LLMs), especificamente o Google Gemini 1.5 Pro, podem superar barreiras na manipulação de dados não estruturados de quaisquer poluentes atmosféricos, desde a extração até a modelagem preditiva. A metodologia foi validada utilizando o histórico de emissões (2001–2023) do Instituto Água e Terra (IAT-PR) e Fundação Estadual de Proteção Ambiental (FEPAM-RS), com Dióxido de Enxofre (SO₂) como estudo de caso. O diferencial metodológico reside na utilização da IA como ecossistema integrado em três etapas: (1) extração automatizada de séries históricas em relatórios PDF mediante engenharia de prompts especializados, explorando a capacidade multimodal do modelo; (2) estruturação, padronização de unidades e análise estatística via geração e execução de código Python em ambiente isolado, garantindo precisão determinística e auditável aos cálculos para mitigar alucinações matemáticas do modelo; (3) implementação de modelos preditivos LSTM (Long Short-Term Memory - Memória de Curto-Longo Prazo) para projeção de cenários futuros. A construção dos prompts seguiu estrutura rigorosa com quatro componentes: objetivo (definição da tarefa), contextualização (domínio ambiental), fonte (identificação dos documentos) e expectativa (formato de saída). Essa arquitetura garantiu reprodutibilidade e estabeleceu protocolo de validação humana em todas as etapas, consolidando a IA como amplificadora, e não substituta, do julgamento técnico. Os resultados demonstram que a metodologia acelera significativamente a transformação de dados em conhecimento aplicável, reduzindo em 78% o tempo de processamento. No estudo com SO₂, identificou-se redução de 35% nas concentrações no Paraná entre 2001 e 2023, mantendo-se em níveis seguros. Para o Rio Grande do Sul, o modelo preditivo (acurácia de 92%) atuou como um sistema de alerta precoce: embora projete conformidade inicial com a

Resolução CONAMA nº 506/2024, revelou uma tendência de estagnação nas reduções que coloca o estado em risco de não cumprimento do Padrão Final e da Agenda 2030 a longo prazo, demandando novas medidas de mitigação. Conclui-se que a metodologia é universalmente aplicável a quaisquer parâmetros de qualidade do ar. Além de transformar relatórios estáticos em instrumentos dinâmicos para a governança ambiental baseada em evidências, o estudo possui fortes implicações pedagógicas: ao eliminar a necessidade de domínio prévio em programação, democratiza o acesso a técnicas complexas de ciência de dados para gestores públicos, professores e estudantes.

Palavras-chave: Inteligência Artificial Generativa; Engenharia de Prompts; Qualidade do Ar; Monitoramento Ambiental.