

# Predição da concentração de poluentes atmosféricos em São Paulo

## Introdução

A deterioração da qualidade do ar é um problema ambiental recorrente na região metropolitana de São Paulo. Em especial, os poluentes  $MP_{10}$  (material particulado inalável) e  $O_3$  (ozônio) frequentemente atingem concentrações acima do recomendado pela OMS (Organização Mundial da Saúde). Enquanto o  $MP_{10}$  é majoritariamente emitido por processos de combustão, o  $O_3$  é um poluente produzido na atmosfera a partir de reações fotoquímicas. A concentração de um poluente na atmosfera depende do balanço entre fontes emissoras, como veículos, indústrias e produção química, de processos de remoção, como a deposição pela ação da chuva e a dispersão pelo vento, e de reações químicas. Quando a magnitude das fontes supera a magnitude dos processos de remoção, ocorre acúmulo do poluente na atmosfera, com aumento da sua concentração. Mudanças nos padrões de emissão podem ocorrer devido à aplicação de políticas públicas e à variabilidade das condições econômicas. Mudanças nas intensidades de remoção podem ocorrer devido a mudanças no clima devido a fatores globais e regionais e devido a eventos extremos de seca e de precipitação.

A proposta deste exercício é prever a concentração do poluente  $O_3$  ao longo do ano de 2022. São fornecidas séries temporais de dados diários de concentração de poluentes monitorados pela CETESB no Parque do Ibirapuera em São Paulo, entre 2005 e 2022. Também serão fornecidos dados de variáveis meteorológicas de superfície monitorados pela estação meteorológica do IAG/USP em São Paulo. Como variável alvo, adotar o poluente  $O_3$ . Como variáveis preditoras, estão disponíveis as séries temporais de 13 atributos (features) que incluem a concentração de poluentes e variáveis meteorológicas. Os dados estão disponíveis [neste link](#).

Para resolver este problema, recomenda-se cumprir as seguintes etapas:

- Análise exploratória de dados: realizar análises descritivas e visualizações para entender a distribuição das variáveis e identificar possíveis padrões.
- Seleção de atributos (features) relevantes para o problema.
- Pré-processamento de dados.
- Adotar uma estratégia de validação.
- Adotar métricas de desempenho adequadas.
- Aplicar duas técnicas de aprendizado supervisionado para resolver este problema de regressão.
- Avaliar e comparar o desempenho dos diferentes modelos.

## Objetivos

Desenvolver modelos de regressão supervisionados que reproduzam a variabilidade temporal do poluente  $O_3$  na área de estudo com base em variáveis meteorológicas e concentração de outros poluentes.

## Entregáveis

1. Relatório técnico detalhado contendo a descrição dos dados, metodologias utilizadas, resultados das análises, interpretação dos insights obtidos e conclusões.
2. Apresentação oral do projeto destacando os principais resultados, desafios enfrentados e possíveis aplicações práticas dos insights obtidos.

## Recursos Necessários

Dados diários de concentração de poluentes (CETESB) e dados diários de variáveis meteorológicas (IAG/USP), entre 2005 e 2022. O conjunto de dados possui 13 atributos (features), além da variável alvo ( $O_3$ ). Os dados estão disponíveis [neste link](#).

## Equipe

O projeto será realizado por uma equipe de alunos da disciplina de Inteligência Artificial, com orientação e supervisão dos professores responsáveis pela disciplina.