Milestone 10: Location-Aware PM2.5 Prediction Model

# Objective

To improve prediction accuracy by training a linear regression model that incorporates geographic location as an additional input, enabling location-aware PM2.5 forecasts.

# Data Preparation

- Filtered dataset to only include entries for 'Fine particles (PM 2.5)'.

- Parsed and extracted date features: month, year.

- Engineered cyclical time features: sin(month), cos(month).

- Encoded each unique 'Geo Join ID' as a numeric location code.

- Created lag features: pm25\_lag\_1, pm25\_lag\_2, pm25\_lag\_3.

# Model Training

- Trained a new Linear Regression model using the following features:

* • pm25\_lag\_1, pm25\_lag\_2, pm25\_lag\_3
* • month, sin(month), cos(month)
* • location\_code

- Model was saved to 'ai/models/geo\_model.joblib'.

# API Integration

- Created new endpoint: POST /v1/predict/geo.

- Automatically computes sin\_month and cos\_month from the input month.

- Requires: pm25\_lag\_1, pm25\_lag\_2, pm25\_lag\_3, month, location\_code.

# Testing

- Added automated unit tests in 'test\_predict\_geo.py' to validate the endpoint.

- Verified correct response structure and error handling for invalid input.

# Example Usage

curl -X POST https://dscp-kskv.onrender.com/v1/predict/geo \  
 -H "Content-Type: application/json" \  
 -H "x-api-key: your-api-key" \  
 -d '{  
 "pm25\_lag\_1": 8.7,  
 "pm25\_lag\_2": 8.4,  
 "pm25\_lag\_3": 8.1,  
 "month": 5,  
 "location\_code": 42  
 }'

# Conclusion

This milestone added geographic context to the PM2.5 prediction model, improving accuracy by incorporating location-specific trends. The new model is fully integrated into the API and has been tested for robustness and reliability.