Islamabad Air Quality Predictor

Air Quality Forecasting

By: Ayema Haris

Introduction

Air pollution is a critical health concern for residents of Islamabad, Pakistan. During winter months, the city frequently experiences hazardous air quality levels that pose serious risks to public health, particularly for children, elderly individuals, and people with respiratory conditions.

The Islamabad Air Quality Predictor is a machine learning-powered system that forecasts air quality for the next three days, enabling citizens to make informed decisions about their daily activities. Whether it's deciding to exercise outdoors, taking children to the park, or preparing medication in advance, this system provides the crucial information needed to protect health and wellbeing.

This project demonstrates how technology can be used to address real-world public health challenges, making critical environmental data accessible and actionable for everyone.

The Problem We're Solving

Why Air Quality Matters

Poor air quality has severe health impacts. PM2.5 particulates—tiny particles smaller than 2.5 micrometers—can penetrate deep into lungs and even enter the bloodstream, causing respiratory diseases, heart problems, and worsening conditions like asthma.

The World Health Organization recommends PM2.5 levels stay below 15 µg/m³ for a 24-hour average. Islamabad regularly exceeds this guideline, especially during winter when temperature inversions trap pollutants close to the ground.

The Gap: Lack of Advance Warning

While current air quality data is available, residents lack advance warning systems. Without knowing tomorrow's air quality, people cannot:

- Plan outdoor activities safely
- Prepare necessary medications in advance
- Adjust commuting schedules to avoid peak pollution hours
- Make informed decisions about children's school activities
- Arrange indoor alternatives for vulnerable family members

This project addresses this gap by providing accurate three-day forecasts, giving residents the time they need to plan and protect their health.

How This System Helps

What the System Does

The Islamabad Air Quality Predictor is a fully automated system that:

1. Collects hourly air quality data from reliable sources
2. Uses advanced machine learning to predict PM2.5 levels for the next 72 hours
3. Converts predictions into easy-to-understand Air Quality Index (AQI) categories
4. Updates forecasts daily with the latest predictions
5. Displays information through an accessible web dashboard

The system is designed to be completely hands-off—it runs automatically every day, ensuring forecasts are always current without any manual intervention.

Understanding AQI Categories

The Air Quality Index makes complex pollution data simple to understand. Here's what each category means:

AQI Range	Category	Health Impact
0-50	Good	Air quality is satisfactory. Safe for everyone.
51-100	Moderate	Acceptable for most. Unusually sensitive people should limit prolonged outdoor exertion.
101-150	Unhealthy for Sensitive Groups	Children, elderly, and people with heart/lung disease should limit outdoor activities.
151-200	Unhealthy	Everyone may experience health effects. Sensitive groups should avoid outdoor exertion.
201-300	Very Unhealthy	Health alert: everyone should minimize outdoor exposure.
301+	Hazardous	Health emergency: everyone should avoid all outdoor exposure.

Real-World Benefits

This system provides tangible benefits to different groups:

For Families with Children

Parents can check the three-day forecast to plan outdoor activities when air quality is safe. If unhealthy conditions are predicted, they can arrange indoor alternatives like visits to museums or malls, protecting children's developing lungs from harmful pollution.

For Individuals with Health Conditions

People with asthma, heart disease, or respiratory conditions can prepare medication in advance and adjust their schedules. Knowing three days ahead gives time to consult doctors and stock up on necessary treatments before pollution peaks.

For Fitness Enthusiasts

Runners, cyclists, and outdoor fitness enthusiasts can schedule workouts for days with better air quality, or shift to indoor gyms when pollution is high. This prevents the health risks of heavy breathing during exercise in polluted air.

For Schools and Organizations

Educational institutions and community organizations can plan outdoor events when conditions are safe, and reschedule or move activities indoors when air quality is predicted to be poor.

How the System Works

The system operates through four automated stages:

1. Data Collection (Every Hour)

The system automatically collects air quality measurements every hour from the Open-Meteo API, which provides data for Islamabad's coordinates. This includes PM2.5, PM10, and other pollutants, along with time information that helps the model understand daily and seasonal patterns.

2. Model Training (Daily)

Every day, the system trains machine learning models using all historical data. It compares three different prediction methods and automatically selects the most accurate one. The model learns patterns like rush hour pollution spikes and seasonal variations, continuously improving its accuracy.

3. Forecast Generation (Daily)

Using the best model, the system generates hour-by-hour predictions for the next 72 hours. It fetches weather and air quality forecasts, runs them through the trained model, and converts PM2.5 predictions into user-friendly AQI categories with color coding.

4. Dashboard Display (Always Available)

The web dashboard displays the forecast in an easy-to-read format. Users see today's air quality, a three-day outlook with color-coded categories, a 72-hour trend chart, and specific health advisories based on the predicted AQI level. The dashboard updates automatically every day without requiring any manual work.

Accuracy and Reliability

The system's predictions are highly accurate. The machine learning model achieves 98.89% accuracy, with an average prediction error of only 2.3 µg/m³. For context, this means if the actual PM2.5 level is 50 µg/m³ (moderately unhealthy), the system's prediction will typically be within 2-3 µg/m³ of that value—well within the range needed for reliable health guidance.

Data collection happens every hour, training and predictions run automatically every day, and the dashboard stays continuously updated. This reliability means residents can trust that forecasts are always current and based on the latest information.

Challenges Faced and How To Overcome Them

Building this system involved solving several technical challenges. Each challenge taught valuable lessons about building reliable, automated systems.

Challenge 1: Software Compatibility Issues

The Problem:

When first setting up the machine learning libraries, installation failed because the software versions weren't compatible with the operating system.

The Solution:

Updated to newer versions of the machine learning libraries that supported the modern Python version, ensuring all components worked together smoothly.

Challenge 2: Connection to Cloud Storage

The Problem:

The system couldn't connect to Hopsworks (the cloud platform storing our data and models). Errors indicated authentication problems and incorrect server addresses.

The Solution:

Fixed three issues: corrected the server address to point to the European instance, obtained the proper API key from the project settings, and ensured configuration files used the correct text encoding.

Challenge 3: Automation Permissions

The Problem:

The automated system on GitHub tried to save updated predictions but was denied permission, preventing the dashboard from showing the latest forecasts.

The Solution:

Enabled write permissions in GitHub's repository settings, allowing the automated workflows to save prediction files back to the repository.

Challenge 4: Dashboard Deployment

The Problem:

When deploying the web dashboard to Streamlit Cloud, it failed to build because the cloud service used a newer Python version that wasn't compatible with our data processing libraries.

The Solution:

Created a configuration file specifying Python version 3.11, forcing the cloud service to use a compatible version that worked with all our libraries.

Conclusion

The Islamabad Air Quality Predictor demonstrates how technology can address real public health challenges. By providing accurate three-day forecasts, this system empowers residents to make informed decisions that protect their health and the health of their families.

The project achieves its goals through:

- High prediction accuracy (98.89%) ensuring reliable forecasts
- Full automation requiring no manual intervention
- Accessible web dashboard available to everyone
- Daily updates ensuring information stays current
- Clear health guidance based on AQI categories

Most importantly, this system serves a genuine community need. Whether helping parents plan safe outdoor activities, enabling individuals with health conditions to prepare in advance, or allowing fitness enthusiasts to exercise when air is cleaner, the Islamabad Air Quality Predictor puts valuable information in the hands of citizens who need it most.

As air quality continues to be a critical health concern in Islamabad and cities worldwide, tools like this demonstrate the potential for technology to improve public health outcomes and quality of life.