Enhancing Air Quality Prediction and Public Awareness in Asunción

Fernanda Carlés ♠ Diego Stalder ♣ Alvaro Machuca† Soledad Benítez⋆ Carolina Recalde♢ Luis Bernal♡

Faculty of Engineering - National University of Asunción

♠fernanda.carles@gmail.com ♣dstalder@ing.una.py †alvaro@machucarecal.de

⋆solbenitezca@gmail.com ♢nrecalde@ing.una.py ♡lbernal@ing.una.py

Introduction

Air pollution poses significant health risks, causing millions of deaths annually worldwide. In Asuncion, Paraguay, local activities and seasonal forest fires elevate pollutants, particularly PM2.5 and PM10. To address this, the Faculty of Engineering at the National University of Asunción (UNA) established a network of air quality monitoring stations in 2019. Using this data, we developed machine learning models predicting Air Quality Index levels in Asunción, achieving 91% accuracy for 6-hour forecasts and 86% for 12-hour forecasts. To make these forecasts accessible, we are developing an open-source, user-friendly web application with real-time air quality information and alerts.

1 Methodology

Predictor Modeling and Data: In an earlier phase of this project, we employed XGBoost with a 2-day regression window and a 5-minute data frequency to build individual prediction models for each station on the network, focusing solely on pollution and temporal data as features [1].

Building on the findings from the previous study, our current work employs LightGBM[3] with a more refined 1-day regression window and 1-hour data frequency. Importantly, we have expanded the feature set to include not only pollution data but also weather variables and statistical features, and trained generalized models that can handle predictions for any station within the network.

Data	Features
Pollution	pm2_5, pm10, pm1, **aqi_2_5 (TARGET)**
Weather	temperature, humidity, wind_speed, wind_dir
Statistical	pm2_5_6h_max, pm2_5_6h_skew, pm2_5_6h_std, aqi_24h_max, aqi_24h_std
Time	hour, day_of_week

Table 1: Features used for training AQI forecasting models.

Another key aspect of these new models is the correction of PM2.5 measurements for humidity by using the equation provided by Crilley et. al[2]:

$$\frac{m}{m_o} = 1 + \frac{\frac{\rho_w}{\rho_p}\kappa}{-1 + \frac{1}{a_w}} \tag{1}$$

Where $\frac{m}{m_o}$ is the ratio of the OPC-N2 measurements to the reference instrument, ρ_w and ρ_p are the density of water and dry particles, a_w is the ambient humidity and κ a regional hygroscopicity parameter.

System Architecture and Data Pipeline Design: The web application's backend consists of two main components. The first is an ETL orchestration pipeline that performs hourly extractions and transformations for local pollution and weather data from FIUNA, Airnow and Meteostat, and produces forecasts using the ML models. This system is connected to Django REST API system connected to a data visualization frontend. The relation system for the database is built using a medallion architecture schema.

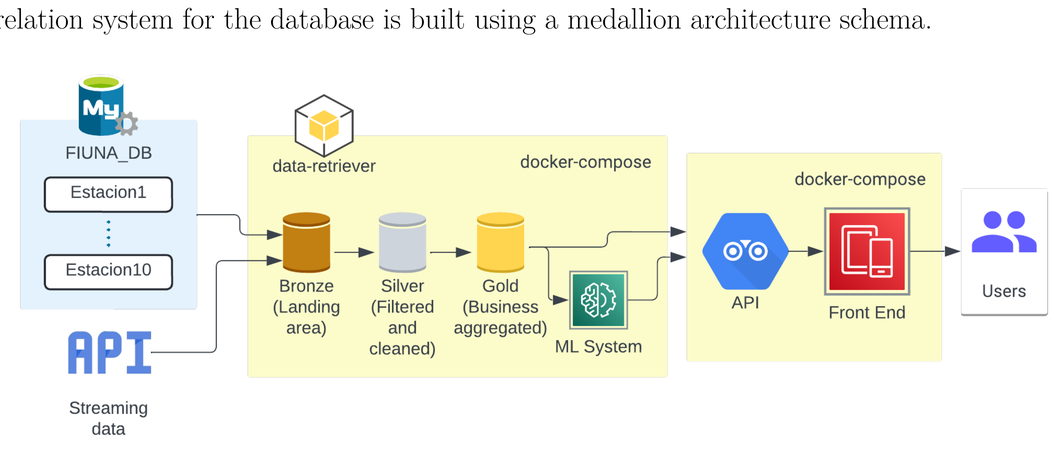

Figure 1: System Architecture

To ensure data consistency for the FIUNA readings, the system is built to perform monthly remote sensor calibrations for each station, using measurements from the US Embassy in Asunción as a pattern station. Pollution readings are also treated for humidity correction by using equation (1).

UX Research and Interface Design: The UX Design process began with workshops by a UX Research team to clarify the prediction system's purpose, target audience, and user benefits. Insights from these workshops informed initial visualizations of the web application. These visualizations, along with a structured questionnaire, were used in interviews with two user groups: User Type A (general public) and User Type B (scientific community). Feedback from these interviews shaped the final UI design.

2 Results

The final web application features an interactive map for visualizing real time air quality data. Users can select map markers to access current AQI values and pollution forecasts for specific city regions. Additionally, the app provides health information and recommendations based on the local air quality.

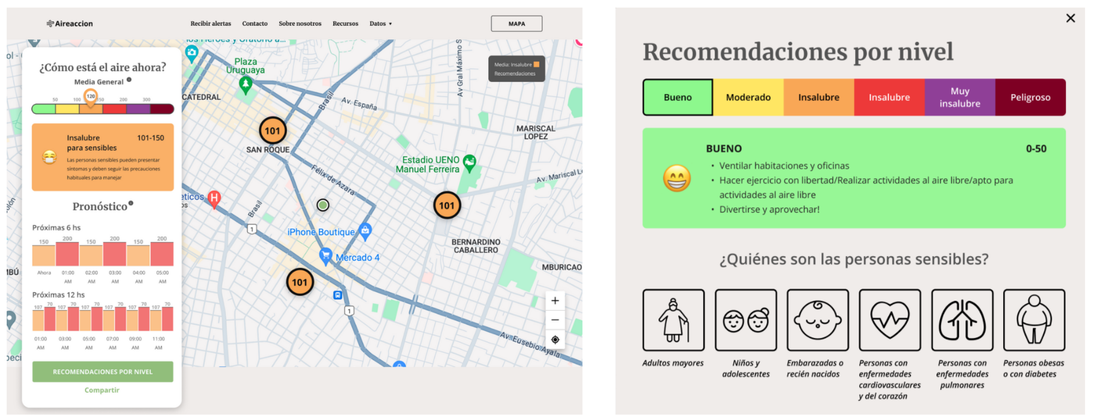

Figure 2: **Web Application Main Page:** Current AQI values, 6 and 12 hour forecasts and health recommendations.

The new prediction models, trained with historical data since March 2020 and tested using data from April 2023 to April 2024, are able to provide hourly 6 and 12 hour horizon forecasts for any specific monitoring city region, and a general AQI forecast for the entire Gran Asunción area.

Horizon	RMSE	MAPE	MAE	R^2
6 hours	5.17	10.51	3.11	0.94
12 hours	9.33	18.07	5.98	0.83

Table 2: Evaluation metrics for air quality prediction models using LightGBM. The values for each metric are the average RMSE, MAPE, MAE and R^2 values for every station in the network.

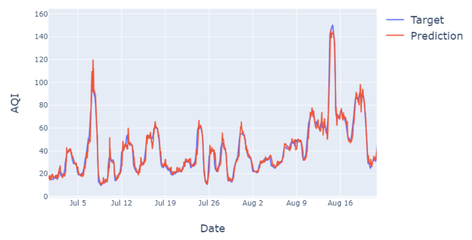

Figure 3: AQI vs Forecast for the 6-hour horizon model. Station 2.

3 Conclusions

This study successfully demonstrates the use of LightGBM models for predicting air quality in Asunción, Paraguay, achieving R^2 scores of 94% and 83% for 6 and 12-hour forecasts respectively. By integrating weather and pollution data with advanced feature engineering, we provide a reliable tool for public health and environmental monitoring.

Our user-friendly web application makes these predictions accessible to both the general public and the scientific community, offering real-time air quality data and health recommendations. Future efforts will focus on enhancing model accuracy and expanding the monitoring network, further supporting efforts to reduce the health impacts of air pollution in the region.

References

[1] Fernanda Carlés, Carolina Recalde, and Diego Stalder. Air quality time series forecasting using machine learning algorithms. In *2023 XLIX Latin American Computer Conference (CLEI)*, pages 1–9. IEEE, 2023.

[2] Leigh R Crilley, Marvin Shaw, Ryan Pound, and Louisa J Kramer. Evaluation of a low-cost optical particle counter (alphasense opc-n2) for ambient air monitoring. *Atmospheric Measurement Techniques*, 11(2):709–720, 2018.

[3] Guolin Ke, Qi Meng, and Thomas Finley. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

Acknowledgements

This project is supported by Girls Code and the Mozilla Foundation through the project "AI-Driven Air Quality Forecasting for Asunción, Paraguay". FIUNA's Air Quality Monitoring network is supported by CONACYT.