Air Quality Monitoring and Anomaly Detection System

BY

AI Saturdays Lagos Cohort 8 - TEAM MANDELA

<u>Proposed Project Name</u>: MandelaAir Guard Pro or Mandela AirWatch/Air Sentinel

Introduction

Air quality monitoring is the process of measuring various pollutants present in the air. The quality of the air we breathe is a paramount concern for public health, environmental sustainability and overall well-being. In today's rapidly urbanizing world, air pollution has become a pervasive issue, with serious implications for individuals and communities. A study conducted by the World Bank, titled 'The Cost of Air Pollution in Lagos', estimates that illness and premature deaths due to ambient air pollution caused losses of $2.1 billion in 2018, representing about 2.1% of Lagos State's GDP. In the same year, it caused 11,200 premature deaths, the highest in West Africa. The air we breathe is invisible, yet its impact on our health and environment is undeniable. It affects everything from our respiratory health to the integrity of our ecosystems.

This project, the Air Quality Monitoring and Anomaly Detection System, endeavors to address these concerns by utilizing the power of machine learning to predict air quality trends and detect anomalies in air quality data which can be challenging to monitor in real-time.

Problem Description

This project aims to leverage machine learning techniques and capabilities to detect anomalies on air quality data and predict air quality trends. It aims to address the problem of forecasting air pollution levels and identifying unexpected pollution irregularities and spikes in environmental data. By utilizing datasets aggregated from multiple sources, the proposed system aims to provide an extensive real-time overview of air quality trends. The system would primarily utilize machine learning models to reach accurate predictions of key air quality indices (e.g. AQI) in countries and cities across the world. The project could potentially utilize regression models (e.g. Gradient Boosting) and other approaches like Long Short-Term Memory (LSTMs) networks to make accurate predictions.

The proposed project also places significant emphasis on the concept of anomaly detection. Using relevant machine learning algorithms, we aim to identify deviations from expected air quality patterns (e.g. spikes in pollution levels). To achieve this, we will identify the relevant features that could potentially indicate air quality anomalies and train the pollution data utilizing relevant algorithms (e.g. SVM Isolation Forest) to highlight and understand key relationships and patterns between features. Other key features we could potentially implement with regard to anomaly detection includes Anomaly scoring (data points suggesting potential anomalies) and alert generation (real time alerts when anomalies are detected).

With early warnings and customized alerts, the Air Quality Monitoring and Anomaly Detection System could potentially be relevant for key stakeholders in the public health sector. It also promotes environmental awareness and supports sustainable urban planning based on historical data. Its quick anomaly identification protects communities and demonstrates how technology can be used to address environmental problems, paving the path for a future that is cleaner and more habitable.

Methodology

Several machine learning approaches can be utilized in solving this problem. We can experiment with multiple algorithms, evaluating their performance using appropriate metrics to select the best performing model.

1. **Source for data**:
    a. Historical air quality data with AQI values
    b. Weather data such as temperature, humidity, wind, etc.
    c. Geographical data such as location, latitude, longitude, etc.
2. **Clean data**
3. **Hybrid model approach**:
    a. ML Regression models such as Gradient Boosting Regressor (GBM) or Long Short-Term Memory (LSTM)
    b. Anomaly detection systems such as Isolation forest or One-class SVM
4. **Evaluate Model**: The AQI prediction with metrics such as MAE or RMSE and the anomaly system with metrics such as precision, recall, etc.
5. **Deployment**
6. **Test and Monitor**

Role Assignment

1. Data Sourcing - All team members
2. Data Cleaning - Yetunde Afolabi
3. Model Development I (AQI Model)- Abdul-lateef Asafa
4. Model Development II (Anomaly Detection) - Peter Agida
5. Evaluation - Damilola Akin-Adamu
6. Deployment - Oluwapolore Oyeniji
7. Front-End Development - Undecided
8. Testing - All team members

Conclusion

By embarking on this project, we hope to empower individuals, communities and organizations with the knowledge they need to make informed decisions about their daily activities and policies. We also aim to contribute to the ongoing efforts to combat air pollution, protect public health and ensure a sustainable and clean environment for present and future generations.

Team Mentor

Joscha Cüppers

Team Members

1. Peter Agida
2. Yetunde Afolabi
3. Damilola Akin-Adamu
4. Oluwapolore Oyeniji
5. Abdul-lateef Asafa

References

URLs to some potentially relevant datasets:

- https://ourworldindata.org/air-pollution
- https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset
- https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set
- https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database/2022

URLs to some relevant papers with different ML approaches

- https://www.sciencedirect.com/science/article/pii/S0304380005005259
- https://www.sciencedirect.com/science/article/pii/S1352231099800016
- https://www.sciencedirect.com/science/article/pii/S1309104219304556