

1) Project Name

AI-Powered Air Quality Prediction for Sustainable Cities

2) Overview

This project aims to develop an AI-based model to predict air quality by analyzing historical pollution data. By utilizing machine learning techniques, the model will forecast future air pollution levels, helping policymakers and environmental agencies take proactive measures. This approach contributes to Sustainable Development Goal (SDG) 11: **Sustainable Cities and Communities**, by promoting healthier urban environments.

Air pollution significantly impacts human health and the environment. The ability to predict air quality in advance enables authorities to take preventive actions, such as issuing warnings, adjusting traffic regulations, and implementing environmental policies. This project will explore various machine learning models, including **Linear Regression, Time-Series Forecasting (ARIMA, Prophet, LSTM), and ensemble methods (Random Forest, XGBoost)** to identify the best predictive approach.

3) Background

Air pollution consists of harmful substances such as **PM2.5, PM10, CO, NO2, and SO2**, which can cause respiratory diseases and environmental degradation. Factors such as **wind speed, humidity, and temperature** significantly influence pollution levels, making them crucial variables in air quality forecasting.

Existing forecasting models rely on traditional statistical approaches that may not fully capture the complex, non-linear relationships in pollution data. Machine learning techniques provide an opportunity to improve prediction accuracy by learning patterns from large datasets.

4) Key Objectives / Business Objectives

a) Research Questions:

1. What are the key factors influencing air pollution variations over time?
2. How accurately can machine learning models predict future air quality levels based on historical data?
3. Which model (Linear Regression, ARIMA, LSTM, Random Forest, XGBoost) performs best for air quality forecasting?
4. How can external variables (such as traffic density, weather conditions, and seasonal effects) be incorporated into the prediction model?

5. Can AI-based air quality predictions support decision-making for urban planning and public health policies?

b) Key Steps:

1. **Data Identification and Collection:** Gathering air pollution data from publicly available sources.
 2. **Data Integration and Cleaning:** Handling missing values, normalizing data, and standardizing formats.
 3. **Exploratory Data Analysis (EDA):** Identifying trends, correlations, and seasonal effects.
 4. **Model Development:** Training different machine learning models to forecast air quality.
 5. **Model Evaluation:** Comparing prediction accuracy using metrics such as RMSE and MAE.
 6. **Dashboard Development:** Creating a user-friendly interface for visualizing air quality predictions.
 7. **Reporting and Dissemination of Results:** Summarizing findings and policy recommendations.
-

5) Methods and Workflow

a) Datasets:

The project will use publicly available datasets for air quality prediction, including:

1. **ULUSAL HAVA KALİTESİ İZLEME AĞI (UHKİA)**
(https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew)
 - Government-provided air pollution monitoring data.
2. **Kaggle Air Quality Datasets** (<https://www.kaggle.com/datasets>)
 - Various datasets related to air pollution levels and forecasting models.

b) Data Cleaning and Preprocessing:

1. Handling missing values and removing outliers.
2. Standardizing pollutant concentration units.
3. Creating additional features, such as moving averages, time lags, and meteorological factors.
4. Splitting data into training and testing sets for model validation.

c) Modelling:

1. **Develop Baseline Models:** Implement traditional statistical models such as Linear Regression and ARIMA.
2. **Machine Learning Implementation:**
 - Random Forest and XGBoost for feature importance analysis.
 - LSTM for time-series forecasting.
3. **Hyperparameter Tuning:** Optimize models using Grid Search and Cross-Validation.
4. **Validation and Testing:** Evaluate models using test data and real-world validation.

d) Deliverables:

1. A trained AI model capable of predicting air quality levels.
 2. Performance evaluation comparing different forecasting techniques.
 3. A research paper detailing methodology, findings, and potential applications.
 4. Open-source code repository for future research and replication.
 5. Presentation materials summarizing key insights for policymakers and stakeholders.
-

6) Sustainable Development Goals (SDGs)

This project aligns with **Sustainable Development Goal 11 (Sustainable Cities and Communities)**:

- Enhancing air quality predictions contributes to better urban planning and pollution control.
- Predictions can inform public policies to reduce pollution-related health risks.

7) Timeline and Milestones

The project is planned to span over **one month**, with key milestones including:

- Data collection, cleaning, and exploratory analysis.
- Initial model training and evaluation.
- Model refinement, hyperparameter tuning, and validation.
- Dashboard development, reporting, and final presentation.

Yusuf Mehmet Arslan

Ahmet Sevban Güler

Serhat Keskintürk

Tayfun Ekentok