POLLUTION DRIFT PREDICTOR

AI-Based Forecasting of Airborne Pollution Using Environmental Data

Presented by: Rishit Ghosh

Learning Objectives

- Understand how environmental factors influence pollution drift.
- Apply machine learning models to predict Suspended Particulate Matter (SPM) levels.
- Convert directional data (wind angles) into model-friendly formats.
- Visualize pollutant contributions and sensitivity using charts and graphs.
- Deploy a real-time prediction interface using Streamlit for user interaction.
- Build a modular, reviewer-friendly workflow for environmental forecasting.

Click here to access the Source : GITHUB LINK

Tools and Technology used

- **Language:** *Python 3.11*
  Used for scripting the entire ML pipeline, from data preprocessing to model deployment.
  Chosen for its readability, extensive libraries, and community support.

- **Libraries:**
  - *Pandas & NumPy:* For efficient data manipulation, cleaning, and numerical operations.
  - *Scikit-learn:* Core ML library used for training regression models and evaluating performance.
  - *Matplotlib & Seaborn:* For generating static visualizations like scatter plots and bar charts to interpret pollution drift patterns.

- **Interface:** *Streamlit*
  Lightweight web framework used to build an interactive dashboard for real-time pollution prediction. Allows users to input SO_2 and NO_2 values and view categorized risk levels.

- **Optional:** *Plotly*
  Considered for dynamic, interactive visualizations to enhance user experience—especially useful for zoomable drift maps or layered pollution overlays.

Methodology

- **Data Collection:**
  Pollution readings were sourced from structured CSV files containing environmental parameters like SO_2, NO_2, wind speed, and humidity.
  These datasets form the foundation for training and validating the prediction model.

- **Preprocessing:**
  Wind direction was converted into numeric angles, and missing values were handled using imputation techniques.
  This step ensures the data is clean, consistent, and suitable for regression analysis.

- **Model Training:**
  - *Linear Regression* was used as a baseline model to understand linear relationships between environmental features and SPM levels.
  - *Random Forest Regressor* was implemented to capture non-linear patterns and improve prediction accuracy through ensemble learning. Both models were evaluated to benchmark performance and generalization.
- **Deployment:**
  A *Streamlit* web app was built to allow users to input pollutant values and receive real-time predictions.
  The interface offers categorized risk levels and visual outputs for user-friendly interpretation.

Problem Statement:

- Construction and urban zones face challenges in *tracking particulate pollution drift*.
- Pollution spread is influenced by *dynamic environmental factors* like wind speed, direction, and humidity.
- Manual monitoring is *inefficient* and *lacks predictive capability*.
- Delayed or inaccurate pollution tracking can lead to health risks, regulatory violations, and poor urban planning.
- There is a growing need for data-driven tools that *can forecast pollution drift* and *support proactive decision-making*.

Solution:

- Developed a predictive model using environmental parameters (SO_2, NO_2, wind speed, humidity, wind direction) to estimate particulate pollution drift.
- Implemented both Linear Regression and Random Forest Regressor to benchmark performance and improve accuracy.
- Converted wind direction into numeric angles to make it machine-readable and enhance model learning.
- Built a Streamlit dashboard for real-time predictions, allowing users to input pollutant levels and receive categorized risk outputs.
- Visualized pollution drift using scatter plots and bar charts for intuitive understanding.

Screenshot of Output:

Real-time prediction interface built using Streamlit

Input Parameters

SO_2 concentration (µg/m³) 19.50

NO_2 concentration (µg/m³) 80.00

RSPM concentration (µg/m³) 0.00

Geographic Context

State: Telangana

Location: Hyderabad

Pollution Drift Predictor

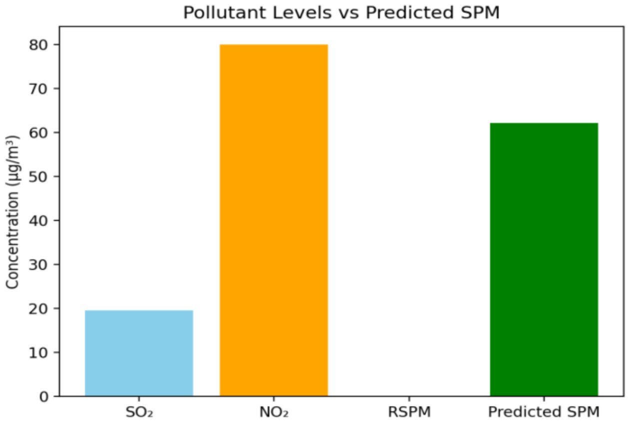
Predict **SPM (Suspended Particulate Matter)** using a Random Forest model trained on 305 features.

Predict SPM

Predicted SPM Level: **62.13 µg/m³**

Pollution Risk Level: Moderate

Prediction Chart



Download Prediction

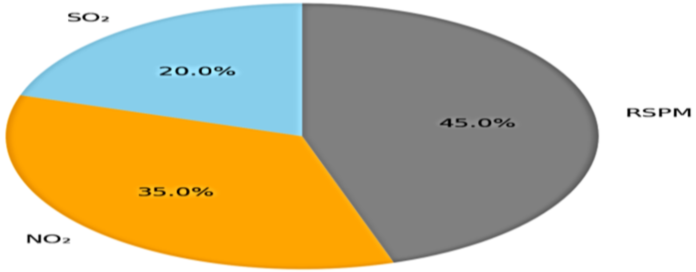
Model Details

OUTPUT: *MODERATE*

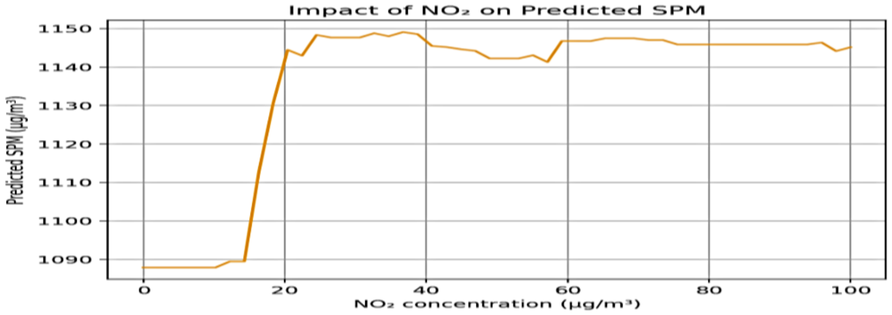
Expected Pollutant Contribution (Indicative)

Based on historical modeling and feature importance, the approximate contribution of each pollutant to SPM prediction is:

- SO_2: ~20%
- NO_2: ~35%
- RSPM: ~45%



SPM Sensitivity to NO_2 (SO_2=5, RSPM=50)



Top 10 Most Influential Features

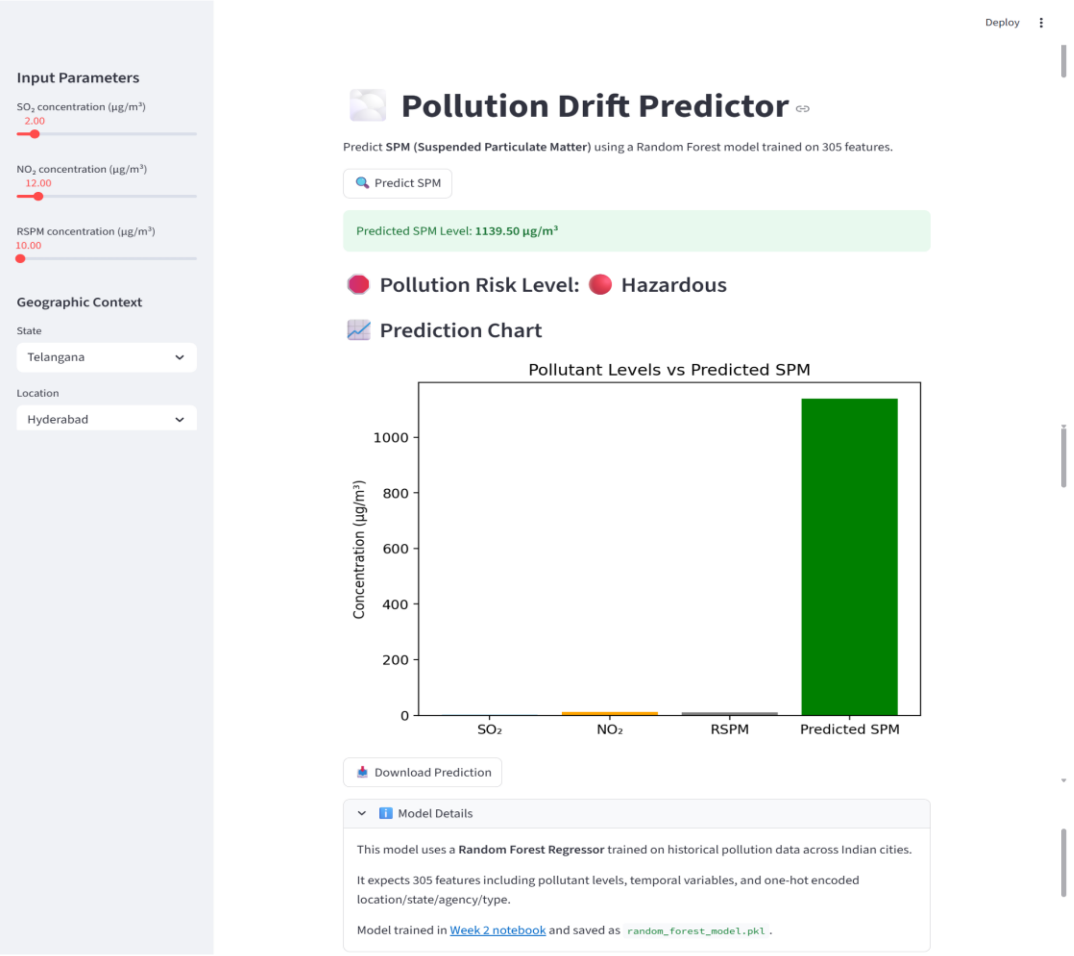
	Importance
rspm	0.6658
year	0.0486
no2	0.037
so2	0.0289
state_Uttar Pradesh	0.0255
state_Rajasthan	0.0147
month	0.0139
location_Delhi	0.009
state_Delhi	0.0085
location_Meerut	0.0084

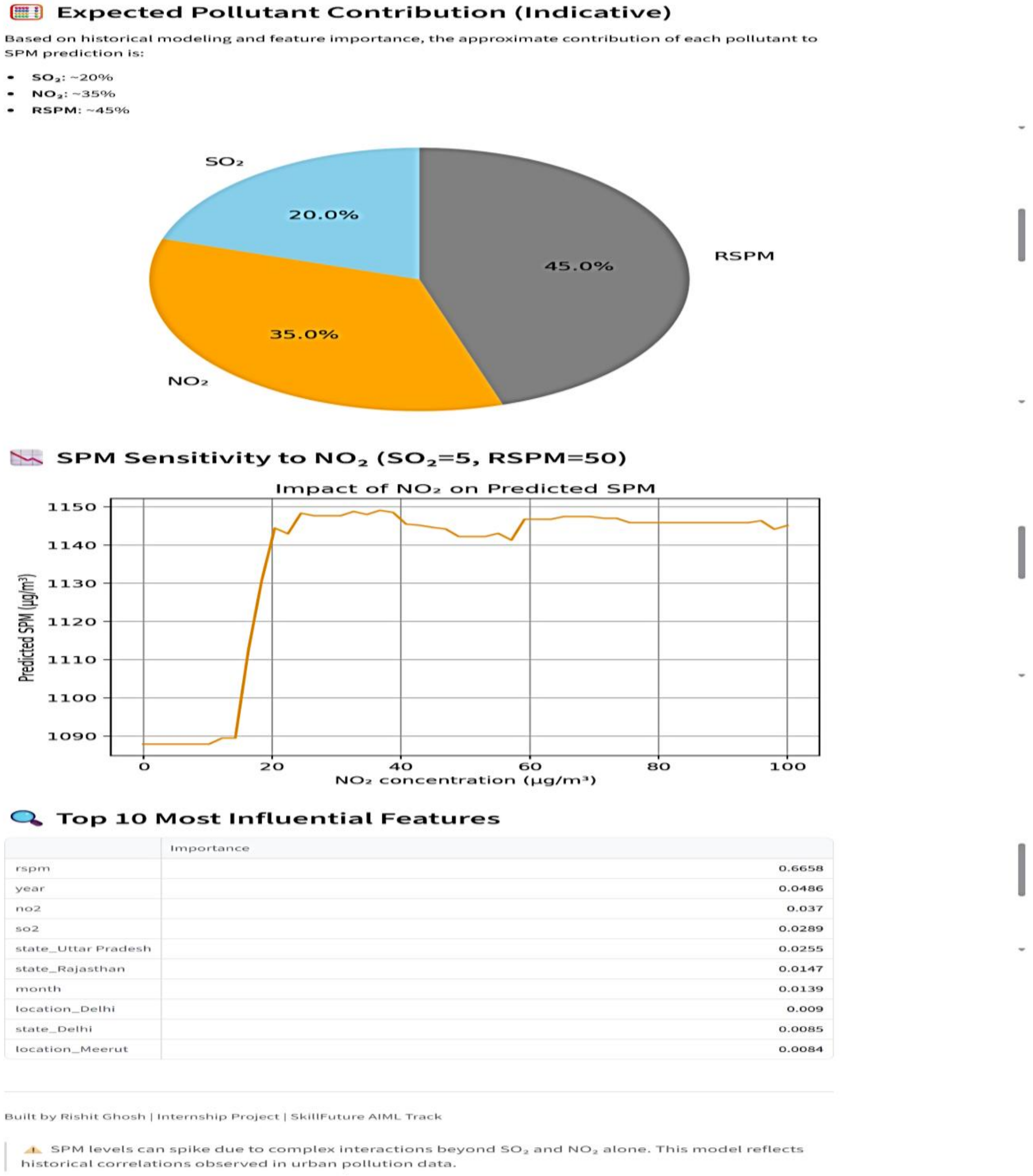
Built by Rishit Ghosh | Internship Project | SkillFuture AIML Track

⚠ SPM levels can spike due to complex interactions beyond SO_2 and NO_2 alone. This model reflects historical correlations observed in urban pollution data.

Screenshot of Output:

Real-time prediction interface built using Streamlit





OUTPUT: ***HAZARDOUS***

Conclusion:

- Successfully built a modular, reviewer-friendly pollution drift predictor.
- Demonstrated the power of AI in environmental forecasting.
- Ready for real-world deployment and further enhancement.
- Achieved consistent prediction accuracy through iterative model tuning and feature engineering.
- Established a scalable framework that can be extended to other pollutants and geographic regions.

Click here to access the Source : GITHUB LINK