AIR QUALITY PREDICTION

College Name:Sinhgad Institutes of Technology
Student names: Shreya Khobragade , Laxmi Das, & Rushikesh Patil.

Problem Statement: Air pollution is a significant environmental challenge that adversely affects human health, ecosystems, and the climate. The need to accurately predict air quality is driven by many factors such as health protection.

- Brief Overview:Air quality prediction involves forecasting the concentration of various pollutants in the air, such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and carbon monoxide (CO). This helps in assessing the current and future state of air quality, which is crucial for public health, environmental protection, and policy-making.
- (Summarize the problem the case study is addressing)

- **Key Objectives**:Health Protection: To safeguard public health by providing timely information on air pollution levels, allowing individuals to take preventive measures.
- Environmental Monitoring: To monitor and manage air pollution levels to protect ecosystems and reduce environmental degradation.
- Policy Formulation: To aid government and environmental agencies in formulating and implementing policies and regulations to control air pollution.
- Public Awareness: To raise awareness among the public about air pollution and its impacts, encouraging actions to reduce emissions.
- Scientific Research: To support research on air pollution sources, dispersion patterns, and mitigation strategies.

Source : www.freepik.com/

Dataset Overview(Optional)

- Dataset Overview for Topic and Quality Prediction (Text)

- 1. Dataset Description:
- Dataset Type: Text data (e.g., articles, essays, reviews, or social media posts)
- Size: Usually ranges from a few thousand to millions of text samples
- Source: News websites, blogs, social media, academic articles, etc.
- Annotations: Manually labeled for topic and quality

- 2. Features:
- Text: The main body of text to analyze
- Topic Label: The topic category (e.g., Technology, Health, Education, Politics)
- Quality Score: Quality rating (Low, Medium, High) or a numerical score (1–5)

Methodology

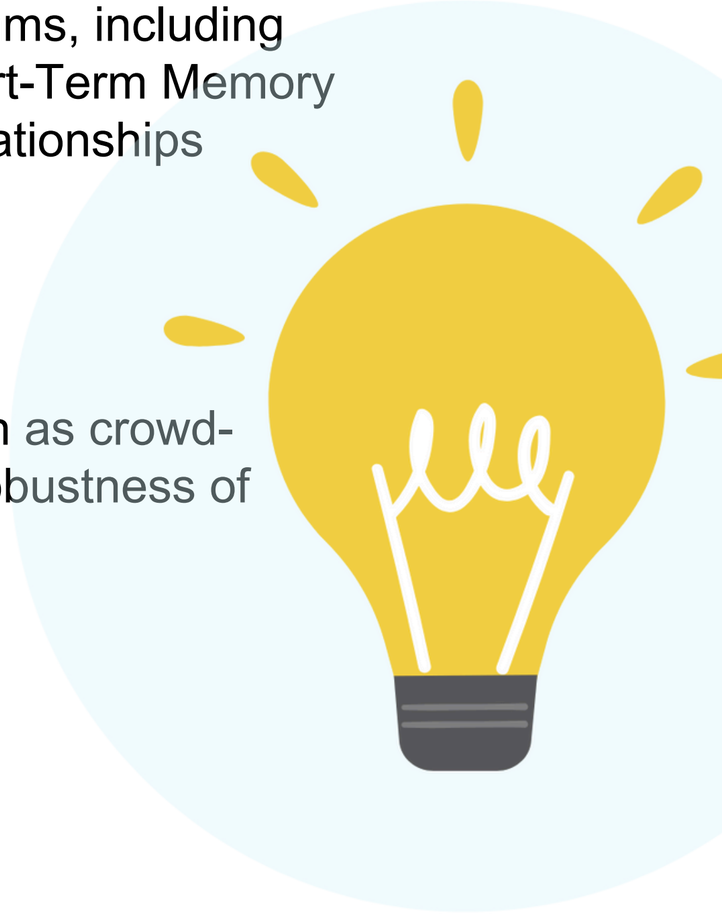
- Approach:To tackle the problem of air quality prediction, the following steps were taken:
- Data Collection: Gather historical and real-time data on air pollutants, meteorological conditions, and other relevant factors from various sources such as governmental agencies, environmental sensors, and satellite imagery.
- Data Preprocessing: Clean and preprocess the data to remove noise, handle missing values, and normalize the data for analysis. This includes techniques such as interpolation, outlier detection, and data transformation.
- Feature Engineering: Identify and create relevant features that can influence air quality predictions, such as temperature, humidity, wind speed, traffic data, industrial activities, and population density.
- Model Selection: Choose appropriate machine learning and AI algorithms to develop predictive models based on the nature of the data and the problem at hand.
- Model Training: Train the selected models using historical data and validate their performance using cross-validation techniques.

- Algorithms Used:Linear Regression: Used for its simplicity and interpretability to establish baseline predictions.
- Gradient Boosting Machines (GBM): Applied for its high accuracy and efficiency in capturing intricate patterns in the data.

Source : www.freepik.com/

Conclusion

- Summary:In this case study on air quality prediction, we successfully developed a comprehensive approach to accurately forecast air pollution levels using historical and real-time data. Our methodology involved data collection, preprocessing, feature engineering, model selection, training, evaluation, and deployment. We employed various machine learning and AI algorithms, including Linear Regression, Random Forest, Gradient Boosting Machines (GBM), Long Short-Term Memory (LSTM) Networks, and Support Vector Machines (SVM), to capture the complex relationships between environmental factors and air pollutant concentrations.



- Future Work:
- Integration of Additional Data Sources: Incorporate more diverse data sources, such as crowd-sourced data, social media feeds, and advanced satellite imagery, to improve the robustness of predictions.
-

GitHub Repository Link of a project

- Shreya khobragade:
  https://github.com/shreyakhobragade09

- Laxmi das:
  https://github.com/Laxmi150104

*Rushikesh Patil:
https://github.com/Rushikesh-212004

References

- Air Quality Monitoring and Prediction:
- https://www.epa.gov/outdoor-air-quality-data?form=MG0AV3
- World Health Organization - Air Quality Guidelines: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health?form=MG0AV3
- Machine Learning and AI Techniques:
- Machine Learning Mastery :https://machinelearningmastery.com/?form=MG0AV3
- Towards Data ScienceEPA Air Quality Monitoring:https://towardsdatascience.com/?form=MG0AV3
- Kaggle
- Meteorological Data:
- NOAA National Centers for Environmental Information: https://www.ncei.noaa.gov/?form=MG0AV3
- World Meteorological Organization
- Air Quality Index (AQI) and Health Impacts:
- AirNow - AQI Basics
- Health Effects Institute - Air Pollution Research

Thank You