**IBM NAAN MUDHALVAN – PHASE 2**

DOMAIN – DATA SCIENCE

AIR QUALITY ANALYSIS & PREDICATION USING DATA SCIENCE

**PROBLEM STATEMENT :**

Air pollution poses a significant threat to public health, environmental sustainability, and overall quality of life in urban areas. The difficulty lies in creating reliable predictive models that make use of both historical and current data, such as pollutant concentrations, weather conditions, traffic patterns, and geographic characteristics, to estimate air quality levels in particular places.

**Solution**

Air pollution is the most considerable environmental health risk in all of Europe. It accounts for mainly cardiovascular and respiratory diseases, causing loss of healthy years of life and premature deaths.

In urban cities, the emissions from the combustion of fossil fuels for various transport modes, industries, and household activities account for the main percentage of emissions emitted into the atmosphere.

Achieving adequate coverage with an air quality monitoring network includes factors such as population density, location, cost, and maintenance life-cycle of measuring devices. Increasing the number of

Monitoring stations is not feasible as per the limited public administration budgets.

There are different methodologies for air quality analysis & predication, and they depend on five types of parameters:

1. Statistical Methods
2. Spatial Analysis
3. Visualization and Communication
4. Air Quality Index (AQI) Calculation

Statistical methods play a crucial role in air quality analysis and prediction using data science.

**1. Descriptive Statistics:**

**Mean, Median, and Mode:** Calculate the average, middle, and most frequent values in pollutant concentrations to understand central tendencies.

**Standard Deviation and Variance:** Measure the dispersion of data points around the mean, providing insights into the variability of air quality parameters.

**Percentiles:** Evaluate the distribution of pollutant levels by calculating percentiles, which can be useful for understanding extreme pollution events.

**2. Regression Analysis:**

**Simple Linear Regression**: Use simple linear regression to model the relationship between one pollutant and a single predictor variable (e.g., temperature or wind speed).

**Multiple Linear Regression:** Extend the analysis to multiple predictors, considering various meteorological and geographical factors simultaneously.

**Non-linear Regression:** Apply non-linear regression models when the relationship between pollutants and predictors is non-linear.

Spatial analysis plays a crucial role in air quality analysis and prediction, allowing researchers and policymakers to understand the spatial patterns of air pollution, identify pollution sources, and make informed decisions regarding urban planning and public health interventions.

**1. Spatial Interpolation:**

**Kriging:** Kriging is a geostatistical interpolation method that estimates pollutant concentrations at unmonitored locations based on the values observed at nearby monitoring stations. It provides predictions with uncertainty estimates, making it valuable for spatial analysis of air quality.

**2. Spatial Variability Analysis:**

**Spatial Autocorrelation:** Spatial autocorrelation analysis helps in identifying patterns in air quality data. It determines whether nearby locations have similar air quality values, indicating spatial clusters of pollution or clean areas.

Visualization and communication are crucial aspects of air quality analysis and prediction using data science. Effectively conveying complex information to various stakeholders, including policymakers, scientists, and the general public, is essential for creating awareness, making informed decisions, and driving positive changes.

**1. Interactive Dashboards:**

Develop interactive dashboards using tools like Tableau, Power BI, or D3.js..

Allow users to filter data based on time, location, pollutants, and other relevant parameters for personalized exploration.

**2 Box Plots and Violin Plots:**

Use box plots and violin plots to display the distribution of pollutant concentrations, providing insights into variability and potential outliers.

Compare pollutant distributions across different locations or time intervals, aiding in comparative analysis.

The Air Quality Index (AQI) is a crucial step in air quality analysis and prediction, as it provides a standardized way to communicate air quality information to the public. The AQI condenses complex air quality data into a single value that represents the overall air quality for a specific location.

1. **Normalization:** Normalize the individual sub-indices to a common scale ranging from 0 to 100. The normalization ensures that each sub-index contributes proportionally to the overall AQI.
2. **Sub-Index Selection:** Identify the maximum sub-index among the individual sub-indices. This maximum sub-index represents the primary pollutant, which is the pollutant with the highest impact on air quality at that location.

**Advantages of air quality analysis and prediction**

Implementing air quality analysis and prediction leads to plenty of new business opportunities. You can do a lot of optimization in:

• Early warning,

• Informing decisions,

• Identifying sources of pollution,

• Targeted intervention,

• Research and analysis.

Let’s discuss these benefits in more depth.

* Early warning

Real-time air quality data can provide early warning of potential health hazards. This allows individuals and organizations to take action to protect themselves and others.

* Informing decisions

It can inform decisions about when and where to take action to reduce pollution. This includes decisions such as temporarily closing a school or factory during a high pollution event.

* Identifying sources of pollution

Used to identify the sources of pollution for example, a specific factory or road,which can inform decisions on how to address the problem.

* Targeted intervention

 Utilized to target interventions in specific areas or neighborhoods, where pollution levels are highest.

* Research and analysis

Useful for research and analysis to better understand the causes of pollution and therefore how to effectively address it.

**CONCLUSION:**

In conclusion, air quality analysis and prediction using advanced data science techniques are crucial for addressing the pressing challenges posed by air pollution. The methods employed, ranging from statistical analyses to machine learning algorithms, allow us to transform complex environmental data into actionable insights