Naan Mudhalvan Project

Air Quality Analysis in Tamil Nadu

Team Members:

~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301

Phase 1: Project Definition and Design Thinking

Project Overview:

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The primary objectives are as follows:

- **Analyze Air Quality Trends**: Explore air quality data to identify trends, seasonal patterns, and changes over time.
- **Identify Pollution Hotspots**: Determine areas within Tamil Nadu that consistently experience high levels of air pollution, enabling targeted interventions.
- **Predict RSPM/PM10 Levels**: Develop a predictive model using SO2 and NO2 levels to estimate RSPM/PM10 levels, aiding in early pollution detection and management.

Tools and Technologies:

- Python
- Visualization Libraries
- Machine Learning Libraries

Design Thinking:

Design thinking is an essential process in problem-solving, emphasizing a user-centric approach. In this phase, we'll outline our approach to understanding and solving the problem.

1. Understanding the Problem:

- ***Define Objectives***: We have identified three main objectives for our project, which are to analyze air quality trends, identify pollution hotspots, and build a predictive model for RSPM/PM10 levels.

- ***Gather Data***: Acquire air quality data from dataset link provided. The dataset includes parameters like SO2, NO2, RSPM/PM10 levels, location information, and date.

- ***Model Development:*** Develop a predictive model using machine learning algorithms, with SO2 and NO2 levels as input features and RSPM/PM10 levels as the target variable.

  Action Plan:
  - Split the data into training and testing sets.

  - Use a simple linear regression model to predict RSPM/PM10 levels based on SO2 and NO2 levels.

  - Evaluate the model's performance using metrics like Mean Absolute Error or R-squared.

2. Ideation and Design:

- ***Analysis Approach***:

  - Data Preprocessing: Clean and prepare the data, remove duplicates and handle missing values and outliers.

  - Exploratory Data Analysis (EDA): Conduct EDA to gain initial insights into the data, identify trends, and visualize air quality patterns.

  - Time Series Analysis: For trend identification, use time series analysis techniques such as decomposition and autocorrelation.

  - Spatial Analysis: Utilize geographical information to map pollution hotspots within Tamil Nadu.

  - Model Development: Develop a predictive model using machine learning algorithms, with SO2 and NO2 levels as input features and RSPM/PM10 levels as the target variable.

- ***Visualization Selection***:

Create a basic dashboard or a set of simple visualizations using Excel, Google Sheets, or a free data visualization tool.

  - Line Charts: Use line charts to visualize temporal trends in air quality data, showing how pollution levels change over time.

  - Heat-maps: Employ heatmaps to display spatial variations in air quality, helping to pinpoint pollution hotspots.

  - Regression Plots: Visualize the relationship between SO2, NO2, and SPM/PM10 levels using regression plots.

3. Prototyping:

- Develop code snippets and prototypes for data preprocessing, EDA, and visualization.

- Experiment with different machine learning algorithms for the predictive model and evaluate their performance.

4. Testing and Feedback:

- Gather feedback to refine the project design and approach.

5. Implementation Plan:

- Develop a detailed project plan outlining tasks, milestones, and timelines.

- Allocate resources and responsibilities among team members.

Time-line:

Week 1: Project Initiation and Data Collection
- Define project objectives and gather initial resources.
- Identify and access data sources.
- Download data and start data preprocessing.

Week 2: Data Preprocessing and Initial Analysis
- Continue data preprocessing, focusing on cleaning and formatting.
- Conduct basic exploratory data analysis (EDA).
- Begin basic data visualization with line charts to show initial trends.

Week 3: In-Depth Analysis and Hotspot Identification
- Perform advanced EDA and time series analysis to identify trends.
- Dive deeper into spatial analysis to locate pollution hotspots.
- Continue refining visualizations and begin feature engineering for modeling.

Week 4: Predictive Modeling and Dashboard Development
- Split data for model training and testing.
- Build and fine-tune the predictive model for RSPM/PM10 levels.

Week 5: Reporting, Feedback, and Presentation
- Summarize project findings in a report.
- Share findings and the initial presentation to gather feedback.

Conclusion

In this phase, we have defined the project objectives, outlined our analysis approach, and identified visualization techniques. The next steps involve data acquisition, data preprocessing, and EDA. Design thinking will guide us through the project, ensuring a user-centric and data-driven approach to solving the problem of air quality analysis and visualization in Tamil Nadu.