Data Analytics Internship Program 2024

Final Project Presentation

Project Name : Predicting & Analyzing AQI (Air quality index) of Delhi
Unique ID : IBM0291
Team name : Avantika Students
College name : Avantika University

Team Members :

1. Gourav Panchal
2. Sanskar Shrivastava
3. Devesh Kumavat
4. Shrey Singh
5. Sagar Manjhi

Introduction :

- **Overview of the Project :** The project analyzes the impact of air pollution on public health in Delhi by correlating pollution levels with health outcomes and predicting future risks using advanced data analysis and machine learning techniques. It aims to provide actionable insights for policymakers to mitigate health risks and improve public health.
- **Objective :** To analyze and predict the impact of air pollution on public health in Delhi, providing actionable insights for mitigating health risks.

Problem Identification:

- **Problem Statement :** Air pollution in Delhi significantly impacts public health, leading to increased respiratory and cardiovascular diseases, hospital admissions, and mortality rates.
- **Significance of the Problem :** Understanding and mitigating the health impacts of air pollution is crucial for improving quality of life, reducing healthcare costs, and protecting vulnerable populations in one of the world's most polluted cities.
- **Relevant SDGs :** The project aligns with Sustainable Development Goals (SDGs) 3 (Good Health and Well-being) and 11 (Sustainable Cities and Communities).

Data Collection :

- **Sources of Data :** I have used a kaggle dataset, from year **2020 to 2023**, Url : https://www.kaggle.com/datasets/deepaksirohiwal/delhi-air-quality
- **Data Description :** Air Pollution Data: Includes concentrations of pollutants like CO, NO, NO2, O3, SO2, PM2.5, PM10, and NH3, recorded at various locations in Delhi.
- Health Data: Covers metrics such as hospital admission rates, incidence of respiratory and cardiovascular diseases, and mortality rates.
- **Relevant SDGs :** Air Pollution Data Collection: Obtained from environmental agencies and government databases through APIs or downloadable datasets.

Data Preprocessing :

- **Data Cleaning Methods:** Remove duplicates, handle outliers, and ensure consistency in data formats.
- **Handling Missing Values:** Use imputation techniques or deletion to address gaps, and interpolation for time-series data.
- **Data Transformation Techniques:** Normalize or standardize data, engineer features for enhanced model performance, and encode categorical variables for machine learning compatibility.

Data Analysis :

- **Analytical Tools and Methods Used:** Python libraries (Pandas, NumPy, Scikit-learn), machine learning models, and data visualization tools for analysis and prediction.
- **Key Findings:** Identified correlations between pollution levels and health outcomes, and predicted future health risks based on current data.
- **Insights Derived:** Provided actionable recommendations for reducing health risks through targeted pollution control measures and policy interventions.

Hypothesis Development :

- **Formulated Hypothesis:** Higher air pollution levels in Delhi are linked to increased respiratory and cardiovascular diseases.
- **Rationale Behind the Hypothesis**: Exposure to high pollutant levels, such as PM2.5 and NO2, is known to negatively impact respiratory and cardiovascular health, as supported by existing research and data.
- **Method for Testing:** Analyze correlations and use regression models to assess the relationship between pollution data and health metrics.

Solution Design :

- **Proposed Solution:** Implement targeted pollution control measures and health interventions based on identified correlations and predictive models to mitigate health risks in Delhi.
- **Implementation Plan:** Develop and deploy data-driven policy recommendations, enhance public health programs, and monitor the effectiveness of interventions.
- **Alignment with SDGs:** Supports SDG 3 (Good Health and Well-being) by improving public health outcomes and SDG 11 (Sustainable Cities and Communities) by promoting cleaner air and sustainable urban environments.

Visualization :

- **Data Visualization Techniques Used:** Utilized charts, graphs, and maps to illustrate pollution levels, health outcomes, and their correlations.
- **Charts, Graphs, Maps Used**: Utilized line charts to show trends over time, bar graphs for comparing pollutant levels and health metrics, and heatmaps to visualize pollution concentrations and health outcomes across different areas of Delhi.
- **Key Visual Insights:** Highlighted high pollution areas linked to increased health risks and visualized trends and patterns in the data to inform policy recommendations.

Conclusion :

- **Summary of Findings:** Identified significant correlations between high pollution levels and increased health issues, revealing critical areas and trends.
- **Impact of Proposed Solution:** Implementing targeted pollution control measures and health interventions is expected to reduce health risks and improve overall public health in Delhi.
- **Future Work:** Expand the analysis to include more cities, refine predictive models, and continuously monitor the effectiveness of implemented solutions for broader impact.

References :

- **Data Sources:** Central Pollution Control Board (CPCB), government health departments, public health surveys, India Meteorological Department (IMD), and weather APIs.
- **Tools and Software Used:** Python (Pandas, NumPy, Scikit-learn), data visualization libraries (Matplotlib, Seaborn), and machine learning frameworks (TensorFlow, Keras).
- **Additional References:** Scientific journals on air pollution and health, government reports on environmental and public health, and relevant research papers on predictive modeling and data analysis.