IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11799

A STUDY ON AI IN AIR QUALITY METRICS

G Praveen[1], Manavendra Singh[2], Dhanush Srinivas[3], Jagatha Venkat Surya[4], Sriraj S R[5], Dr.Sandhya N[6]

Department of Artificial Intelligence and Machine Learning, Dayanada Sagar Academy of Technology[1-5]

Head, Dept. Of AIML, Dayananda Sagar Academy of Technology[6]

Abstract: Air pollution poses a significant threat to the ecosystems. Accurate prediction of pollutant levels is essential for informed decision-making and effective conservation policies. This paper explores the application of linear regression as a predictive tool to estimate air pollutant levels, emphasizing its utility in environmental management. By leveraging historical data and identifying key influencing factors, linear regression can provide transparent, interpretable predictions that aid in setting realistic norms and conservation goals. The study utilizes the "Air Quality Index - New Delhi" dataset to illustrate the application of linear regression in forecasting air quality, demonstrating its potential in proactive environmental monitoring and policy-making.

Keywords: Air Pollution, Linear Regression, Predictive Analysis, Air Quality Index (AQI).

I. INTRODUCTION

1.1 Understanding the Impact of Air Pollution

Pollution, arising from the release of harmful substances into the environment, poses a significant threat to both ecosystems and human health. The accurate prediction of pollutant levels in the air, water, and soil is imperative for informed decision-making and the formulation of effective conservation policies. This literature review delves into the realm of air pollutants, explores various measurement methods, and emphasizes the application of linear regression as a predictive tool to estimate air pollutant levels. By shedding light on how linear regression can bolster the creation of pragmatic norms and feasible conservation policies, this review aims to contribute to the ongoing efforts of authoritative bodies in mitigating environmental pollution.

In statistical analysis, linear regression serves as a foundational technique for modeling the relationship between a scalar response and one or more explanatory variables. Its utility extends to machine learning, where it is employed to predict future variable values based on historical data. With its versatility, linear regression emerges as a potent instrument in foreseeing pollutant levels, thereby facilitating proactive measures to mitigate environmental degradation.

II. RELATED WORKS

2.1 Advancements in Predictive Analysis Models

In the domain of conservation, a plethora of predictive analysis models are utilized to estimate pollutant levels. Noteworthy examples include machine learning techniques such as Random Forest, Support Vector Machines, and neural networks, all of which offer robust predictive capabilities. Moreover, time-series analysis, spatial modeling, and hybrid models combining multiple techniques have proven successful in forecasting pollutant levels. While these existing models provide valuable insights, there remains a pressing need for more accurate and understandable pollution measurement and prediction methodologies. This paper underscores the importance of transparent and user-friendly conservation statistics, empowering individuals to take actionable steps towards achieving realistic pollutant level statistics.

2.2 Technological Innovations in Environmental Monitoring

Cutting-edge technologies, including artificial intelligence (AI), are revolutionizing environmental monitoring and management. Initiatives such as the World Environment Situation Room (WESR) and the International Methane Emissions Observatory (IMEO) leverage AI to analyze complex datasets and facilitate data-driven decisions. These technological advancements underscore the growing importance of harnessing innovation in addressing environmental challenges.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

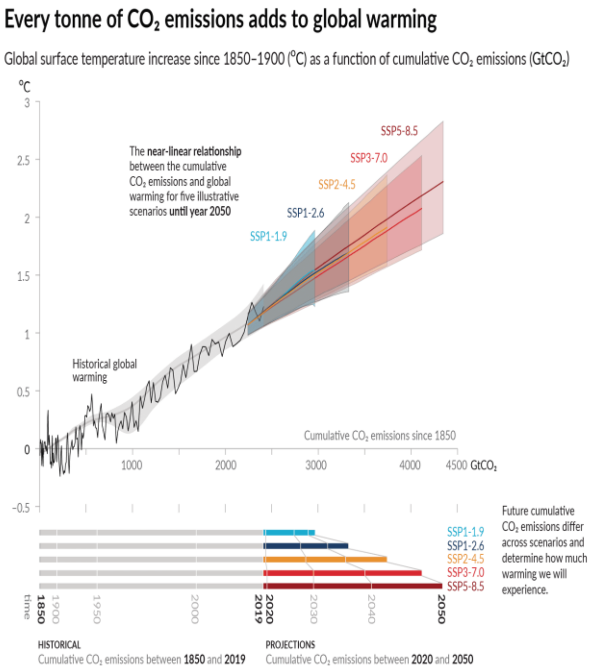
International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 11, Issue 7, July 2024

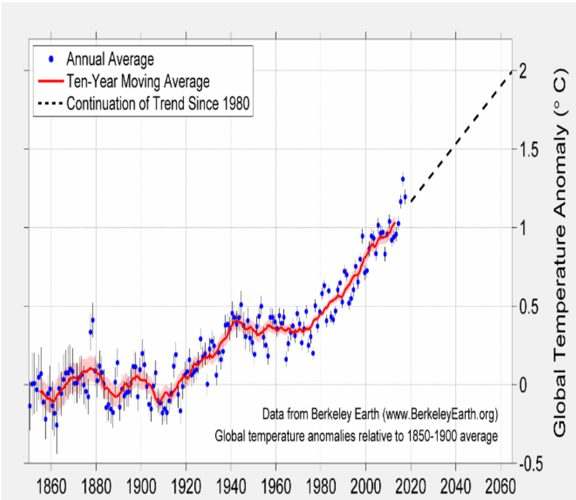
DOI: 10.17148/IARJSET.2024.11799

III. METHODOLOGIES

3.1 Leveraging Linear Regression for Pollution Prediction



Linear regression, a fundamental statistical technique, is increasingly being adopted to predict pollutant levels. This method involves establishing a linear relationship between independent variables (e.g., meteorological data, emission sources) and the dependent variable (pollutant concentration). By analyzing historical data, linear regression models can make accurate predictions and identify influential factors affecting pollutant levels. This approach offers transparency, interpretability, and simplicity, making it a valuable tool for authorities to set realistic norms and conservation goals. Through the application of linear regression, authorities can better understand pollutant trends, identify sources of contamination, and devise targeted strategies for pollution control.



3.2 Utilization of Python Libraries

In implementing linear regression models, Python libraries play a crucial role in data manipulation and visualization. Libraries such as Matplotlib, NumPy, Pandas, and Seaborn provide powerful tools for data analysis, enhancing the accuracy and efficiency of predictive models.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11799

IV. DATASET

In this project, we used the "Air Quality Index - New Delhi" dataset from Kaggle, a widely recognized platform for datasets and data science resources. The dataset provides a comprehensive overview of air quality metrics in New Delhi, a city known for its serious air pollution problems. The data is collected every hour, providing an accurate overview of the changes of air quality at different times of the day and season.

4.1 Data attributes

The dataset contains several important air quality parameters. assessment, including but not limited to:

- Date and time: Time stamps showing the date and time of each recorded data point.

-Concentrations of pollutants: Measurements of major pollutants such as particulate matter matter (PM2.5 and PM10), sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO) and ozone (O3). These concentrations are usually measured in micrograms per cubic meter (μg/m³) or parts per million (ppm), depending on the type of pollutant.

- Air Quality Index (AQI): Calculated indices that indicate overall air. quality are based on pollutant concentrations and are often classified into different levels such as good, moderate, unhealthy, very unhealthy, etc.

- Meteorological data: Other meteorological parameters such as temperature, humidity, wind speed and barometric pressure that can affect the spread of pollution and the dynamics of air quality.

4.2 Data preprocessing

Before analysis, the dataset underwent preprocessing steps such as data cleaning, handling missing values and ensuring data consistency across all variables. Quality assurance processes were applied to validate sensor readings and remove outliers that could distort analytical results.

4.3 Data exploration

Exploratory data exploration (EDA) techniques were used to gain insight into sensor distributions and trends. data set and correlations between variables. Visualization tools such as histograms, time series graphs, correlation matrices and geographic mapping were used to reveal patterns and relationships in the data.

4.4 Data Availability and Updates

It is important to note that the data data The device may have due to air quality monitoring in New Delhi must be updated periodically. According to the following. Continued data collection and integration of new observations will enable more accurate and up-to-date predictive models and environmental assessments.

4.5 Ethical Considerations

Ethical considerations regarding data protection, security and responsible use must be in place. followed when using air quality data. By following appropriate data management practices and obtaining necessary permissions for data use and analysis, ethical research practices and data transparency are ensured. Following this detailed data set and robust analytical methods, our project aims to provide meaningful insights into air pollution dynamics, support informed decision-making processes, and promote sustainable environmental management in urban areas as in New Delhi.

V. CONCLUSION

In conclusion, the accurate prediction of pollutant levels is crucial for effective environmental conservation. Linear regression emerges as a valuable tool due to its simplicity and interpretability. By leveraging historical data and identifying key variables, linear regression can help authoritative bodies make informed decisions, set realistic norms, and establish achievable conservation goals. As we strive towards a cleaner environment, the integration of linear regression into environmental management practices will play a pivotal role in safeguarding our ecosystems and human well-being.

REFERENCES

[1] Matplotlib documentation
[2] NumPy Documentation
[3] Seaborn Documentation
[4] Pandas Documentation
[5] UNEP.org
[6] medium.org