iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60194

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

PradushanCheck - Air Quality Monitoring

Dr. Gresha Bhatia[1], Ashutosh Mishra[2], Muskan Chhabria[3], Nikhil Haswani[4], Vanshika Thakur[5]
Deputy Head of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Abstract: ***PradushanCheck offers a comprehensive approach to air quality monitoring, addressing the critical issue of urban air pollution. Utilizing datasets from major cities including New Delhi, Hyderabad, Kolkata, and Bangalore, we employed advanced techniques such as SMOTE and binning for data preprocessing. Subsequently, we compared the performance of three algorithms—Random Forest Regression, Support Vector Regression, and CatBoost Regression—in predicting air quality indices (AQI) for these cities. Furthermore, we visualized the data on an interactive dashboard using Tableau and integrated the real-time API 'Air Quality Programmatic API' to provide current air pollution status on our website. Our findings underscore the efficacy of PradushanCheck in delivering real-time insights for effective pollution management and public awareness.***
Index Terms: ***Air quality monitoring, Machine learning, Real-time APIs, Air pollution data, SMOTE, Binning techniques, Random Forest, Support Vector Regression, CatBoost Regression.***

I. INTRODUCTION

Concerns over the damaging effects of air pollution on human health and the environment are increasing, which is driving up demand for efficient monitoring and management systems. Heart and respiratory problems are associated with poor air quality, which highlights the significance of accurate monitoring.PradushanCheck quickly offers insights about air quality through the use of cutting-edge technologies, such as real-time API integration and sophisticated data processing. This provides decision-makers with the knowledge they need to effectively combat pollution. The project intends to improve public health and awareness by utilising machine learning to establish a specialised real-time air quality monitoring system. This will nurture a cleaner, safer, and more sustainable urban environment by bringing technology innovation into line with societal needs.

II. MOTIVATION

The urgent and expanding issue of air pollution, which is particularly prevalent in urban areas worldwide, is what motivates us to conduct this study. When one considers the grave health risks and environmental consequences that poor air quality poses, especially in rapidly developing nations like India, the main objective of our initiative becomes evident. With the use of modern machine learning techniques, we hope to revolutionise the field of air quality monitoring. Our dedication to delivering precise, dependable, and fast Air Quality Index (AQI) drives our efforts. Our project's goal is to provide relevant information to stakeholders so they can manage air quality by making informed decisions and taking proactive steps.Our aim is to tackle the pressing need for thorough monitoring of urban air quality by implementing cutting-edge technologies and techniques. This will improve the quality of life and preserve the environment for future generations.

III. LIMITATION EXISTING SYSTEM OR RESEARCH GAP

Existing air quality monitoring systems face several lim-itations such as inferior precision and responsivity, expen-sive installation and maintenance, inaccurate readings, limited spatial coverage, inability to monitor all pollutants, complex equipment, and lack of timeliness for decision-making. These limitations underscore the need for more research on the long-term effects of exposure to air pollutants, relative toxicities from different sources, impacts of ultrafine particulate matter, joint and independent effects of multipollutant exposures, and increased research effort in LMICs (low- and middle-income countries), where exposures are higher but data are scarce. There is also a need for studies on the long-term impacts of air pollution exposure and for complementary means to fill 'gaps' in existing air quality records. These research gaps and limitations underscore the need for innovative solutions like PradushanCheck.

IV. LITERATURE SURVEY

The escalating concerns regarding the detrimental impacts of air pollution on human health and the environment have spurred a growing demand for efficient air quality monitoring and management systems. Research underscores the critical association between poor air quality and various health issues, particularly respiratory and cardiovascular ailments, empha-sizing the urgent need for accurate monitoring and mitigation strategies.

©IJRASET: All Rights are Reserved

In response to these challenges, innovative technologies such as real-time API integration and advanced data processing have emerged to provide rapid insights into air quality con-ditions. Projects like PradushanCheck aim to leverage these cutting-edge tools to enhance decision-making processes and combat pollution effectively. By harnessing machine learning techniques, PradushanCheck endeavors to establish a special-ized real-time air quality monitoring system, thereby fostering cleaner, safer, and more sustainable urban environments.

A. Existing Gaps

Visualization of Air Pollution: Despite the advancements in air quality monitoring technologies, there remains a gap in effectively visualizing air pollution data, particularly at the city level. The lack of comprehensive visualization tools hampers stakeholders' ability to interpret and act upon pollution trends effectively.

Outlier Detection: The presence of outliers or anomalies in air quality data poses a significant challenge to machine learning models used for prediction tasks. These outliers can adversely affect the accuracy of models, leading to suboptimal performance. The need for robust outlier detection techniques is evident to enhance the reliability of air quality forecasting models.

B. Problem Definition

The core objective of this research is to apply machine learning techniques to forecast air pollution levels in urban environments accurately. This entails developing and deploy-ing machine learning models capable of predicting future air quality conditions based on historical and real-time data. The task involves utilizing a range of air pollutant data, including PM2.5, PM10, NO2, SO2, CO, O3, among others, to train predictive models.

The focus on urban environments is paramount, given their unique pollution sources and patterns. Urban areas often expe-rience higher pollution levels due to industrial activities, vehic-ular emissions, and population density. Accurate air pollution forecasts are essential for informing timely decision-making and planning interventions aimed at mitigating pollution levels. By enabling proactive measures, these forecasts have the potential to significantly improve public health outcomes and environmental conditions in urban areas.

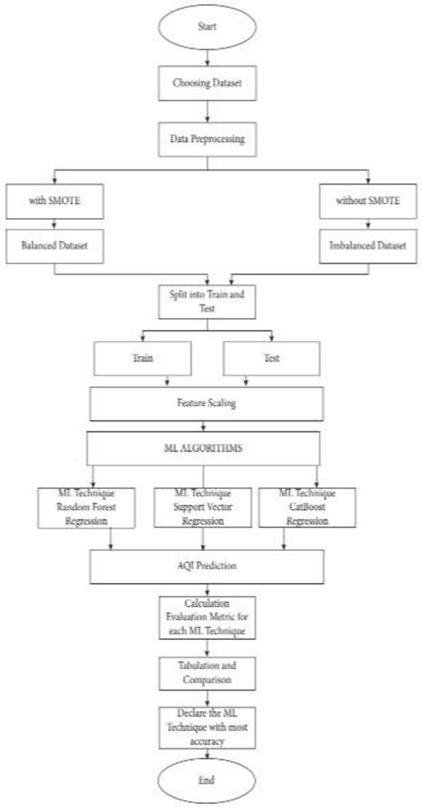


Fig. 1. Flow chart

Regression analysis is a robust method for discovering factors that effect a topic of interest, such as AQI values. It is reliably possible to ascertain which elements are most important, which ones may be disregarded, and how these factors interact by using regression analysis. The variable under prediction is referred to as the dependent variable. In our work, the AQI value is used. Independent variables are those that we presume to have an impact on the dependent variable. This group includes the different weather phenomena..

We can run the model on Python 2.7 (or higher) x64 machine with 8GB RAM, 2.3Ghz intel i7 CPU using the standard visualization and sklearn packages. The majority of work like loading, and handling the null data was implemented using the pandas DataFrame. The dataset was then trained and tested with an approximate ratio of 7:3, ie 4400 train values and 2000 test values. The trained dataset was then fit to predict values using various regression algorithms like Linear Regression, Neural Network Regression, Decision Forest, etc., and compare the results.

V. METHODOLOGY APPLIED

In this project, we aim to predict air quality index (AQI) as the target variable, utilizing various input features includ-ing PM2.5 , PM10, oxides of the Nitrogen. Data collection

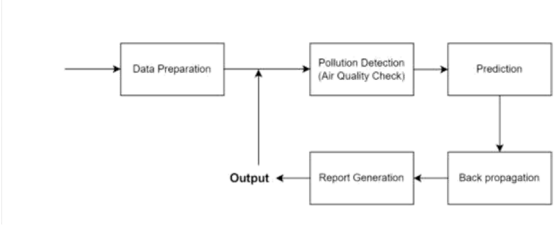


Fig. 2. Block Diagram

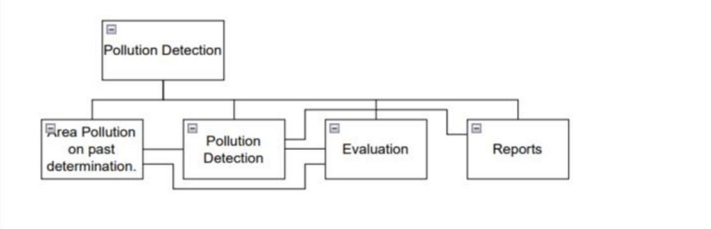


Fig. 3. Modular diagram

involves gathering information from relevant sources such as air quality monitoring stations, satellite images, and weather stations. Preprocessing the data includes handling missing values, normalization, and feature engineering to enhance the model's performance. Different preprocessing methods, including SMOTE and Binning were implemented to address class imbalances within the dataset, resulting in a more bal-anced dataset and improved model performance. Following this, we conducted a comparative analysis of the predictive performance of three distinct algorithms—Random Forest Re-gression, Support Vector Regression, and CatBoost Regres-sion—in forecasting air quality indices (AQI). Additionally, we visually represented the dataset through an interactive dashboard using Tableau software. Moreover, we integrated the 'Air Quality Programmatic API' to provide up-to-date information on air pollution levels on our website. Notably, this API facilitates the retrieval of air pollution data for cities worldwide. Our observations emphasize the effectiveness of PradushanCheck in furnishing real-time insights conducive to efficient pollution management and enhancing public aware-ness regarding air quality.

A. Specific Requirements

1) Software Requirements

a) Language: Python, HTML, CSS, JavaScript

b) IDE: VS Code

c) ML Model Platform: Google Colab

2) Hardware Requirements

a) 2GB Ram

b) Intel Pentium Gold Processor and above / Ryzen Athlon and above

c) 512 Mb Physical Storage and above

B. Algorithm

Random Forest Regression: The Random Forest Regres-sor is an ensemble learning technique that utilizes the decision tree paradigm. During training, it creates several decision trees and combines them to provide outcomes that are more accurate. Evaluate the model's performance on the testing set using appropriate metrics (e.g., Mean Squared Error, R-squared).

C. Hyperparameter Tuning

Fine-tune hyperparameters if needed to improve the model's accuracy. Prediction and Monitoring: Utilize the trained Random Forest Regressor to predict AQI values based on current environmental conditions. Monitor the model's predictions over time to ensure its effectiveness in reflecting changes in air quality.

D. Support Vector Regressor

An example of a supervised learning algorithm is the Sup-port Vector Regressor (SVR), which finds the hyperplane that maximizes the margin between the data and the hyperplane and fits the data the best for regression tasks.With its capacity to manage non-linear connections and adjust to different data distributions, SVR can provide precise AQI monitoring and offer information for well-informed decision-making about the management of air quality in urban settings. It seeks to fit the data inside the tolerance margin while minimizing error.

E. CatBoost Regressor

By training the CatBoost regression model on relevant datasets, incorporating parameters specific to categorical boosting, one can achieve accurate predictions of AQI values. This approach allows the model to adapt to changing con-ditions and improve its forecasting capabilities, contributing to effective air quality monitoring and management in urban environments. CatBoost Regressor employs several optimiza-tions to enhance the training process and improve predictive performance. It makes use of a cutting-edge feature importance calculation approach that sheds light on the relative signifi-cance of various information in forecasting the target variable. .

F. Implementation

Data sourced from New Delhi, Kolkata, Bengaluru, and Hy-derabad, the four prominent cities, is systematically gathered for comprehensive analysis. A visually captivating dashboard is meticulously crafted utilizing the sophisticated data visual-ization capabilities offered by Tableau. This dashboard serves as a powerful tool to present pollution content in an intuitive and insightful manner, facilitating effective interpretation and decision-making processes. In parallel, the website development process employs a robust tech stack comprising HTML, CSS, JavaScript, and React. This amalgamation of technologies ensures the creation of a dynamic and responsive web platform, optimized for seamless user interaction and engagement. To augment the website's functionality and relevance, in-tegration with the Air Quality Programmatic API is imple-mented. This API dynamically retrieves real-time pollution

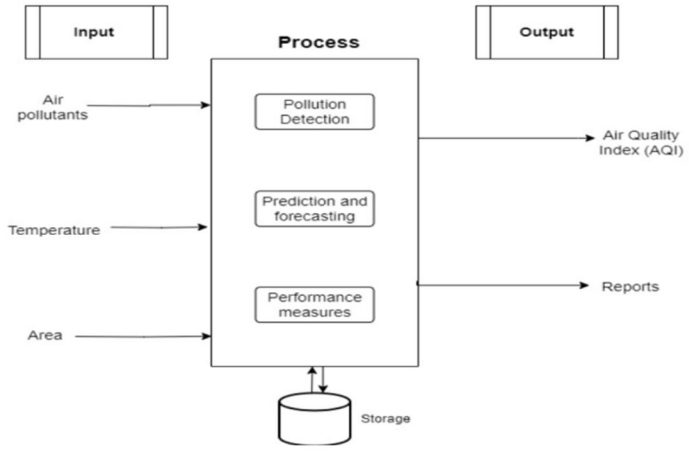


Fig. 4. System Diagram

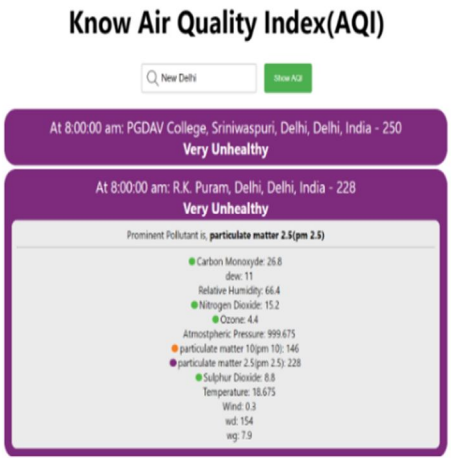


Fig. 6. New Delhi Polluted Areas All Pollution Content



Fig. 5. New Delhi Polluted Areas

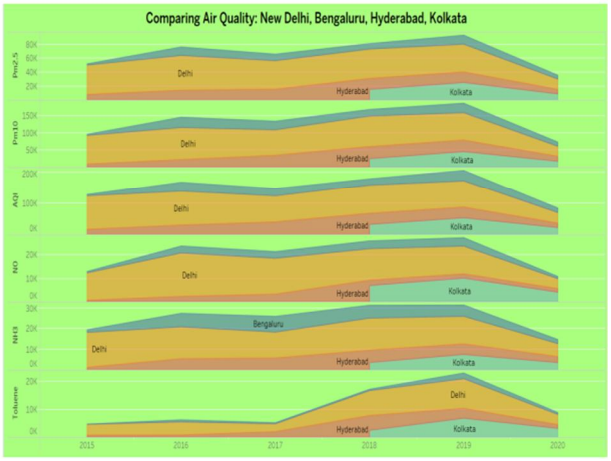


Fig. 7. Comparing Air Quality between 4 Major Cities

data from the aforementioned major cities. Upon user input specifying the city of interest, the website promptly accesses the API to fetch the latest Air Quality Index (AQI) readings and other pertinent pollution metrics. This seamless integration enables users to access up-to-date information on air qual-ity conditions, fostering heightened awareness and informed decision-making regarding environmental health concerns.

G. Results

By providing up-to-the-minute Air Quality Index (AQI) data for a specific area within our city through a user-friendly website, we have empowered residents to make informed decisions about their outdoor activities and take proactive measures to safeguard their well-being. The project's potential extends beyond individual awareness, serving as a valuable resource for policymakers, researchers, and environmental organizations. The collected data can contribute to a better understanding of air quality patterns, aiding in the formulation of evidence-based policies to address pollution concerns and improve overall air quality standards.

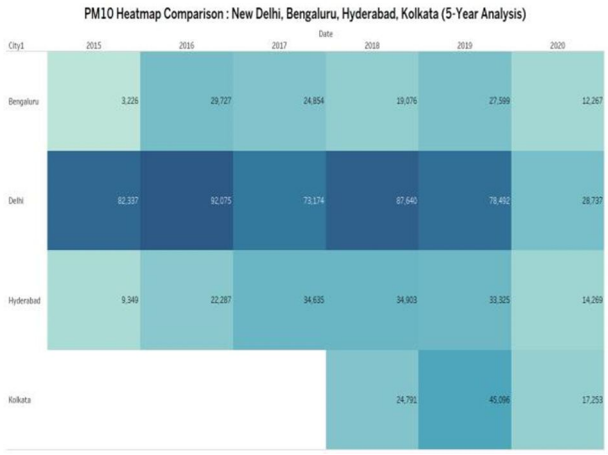


Fig. 8. HeatMap PM10



Fig. 9. HeatMap AQI

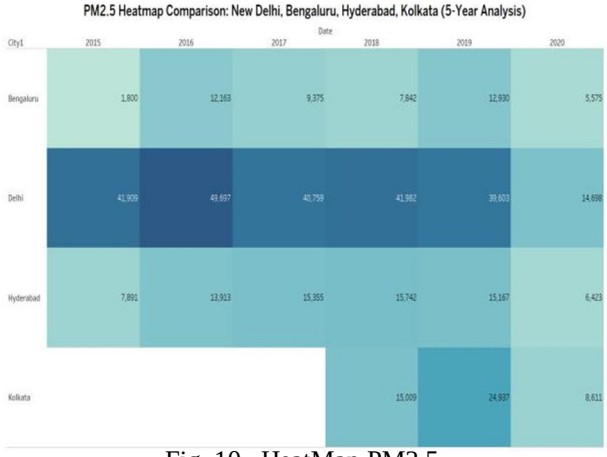


Fig. 10. HeatMap PM2.5

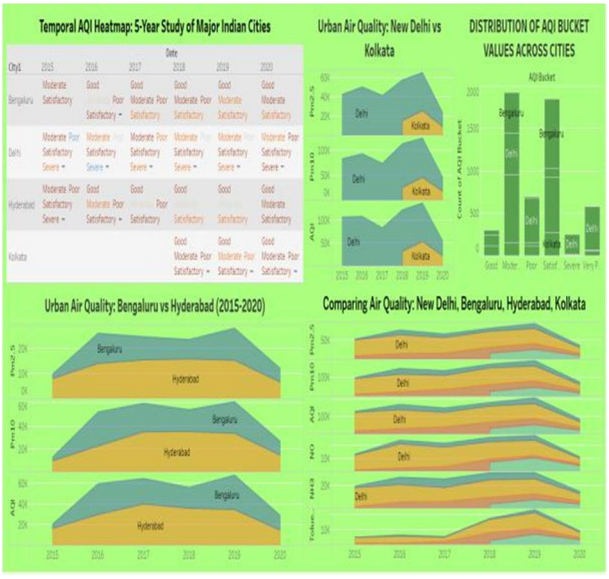


Fig. 11. Dashboard

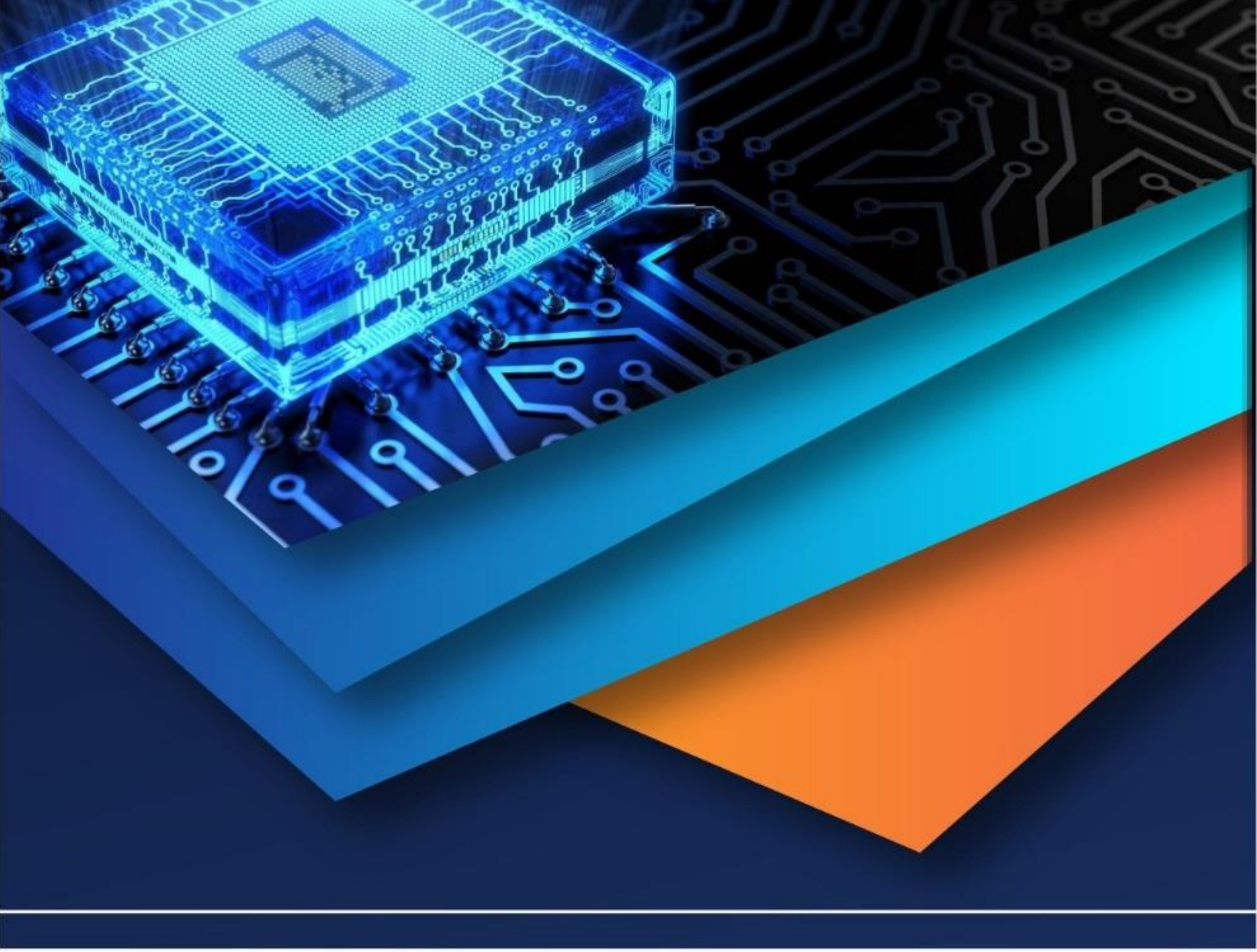
VI. CONCLUSION

The implementation of the CatBoost Regressor in our AQI monitoring website has proven to be a pivotal advancement in predictive modeling and data analysis.Existing systems often calculate the daily Air Quality Index (AQI), but this project aims to predict AQI at a particular area to account for changing weather conditions.Data used for prediction may contain outliers or anomalies, which can adversely affect the accuracy of the machine learning model which have been normalized using SMOTE.

REFERENCES

[1] N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda, Raahil Armaan, B. Valarmathi,and G. Arulkumaran(2023). Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis
[2] A. G. Soundari, J. Gnana, and A. C. Akshaya, “Indian air quality prediction and analysis using machine learning,” International Journal of Applied Engineering Research, vol. 14, p. 11, 2019.
[3] G. Mani, J. K. Viswanadhapalli, and A. A. Stonie, “Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models,” Journal of Engineering Research, vol. 9, 2021.
[4] H. Liu, Q. Li, D. Yu, and Y. Gu, “Air quality index and air pollutant concentration prediction based on machine learning algorithms,” Applied Sciences, vol. 9, p. 4069, 2019.
[5] M. Bansal, “Air quality index prediction of Delhi using LSTM,” Int. J. Emerg. Trends Technol. Comput. Sci, vol. 8, pp. 59–68, 2019.
[6] N. Srinivasa Gupta , Yashvi Mohta, Khyati Heda, Raahil Armaan, Valarmathi and G. Arulkumaran, “Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis” Journal of Environmental and Public Health

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)