iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73722

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Smart Automation for Air Quality Monitoring using ML Techniques

Krupa R K[1], Varadaraj R[2]
MCA, Navkis College of Engineering, Visvesvaraya Technological University

Abstract: Ongoing industrial expansion presents persistent challenges in tracking and evaluating atmospheric quality in metropolitan regions. Assessing environmental air conditions is crucial for establishing contamination levels within specific areas. Atmospheric contamination continues to represent a major worldwide concern affecting both citizens and regulatory authorities. It results in substantial ecological harm, including acidic precipitation and climate change, while creating severe health hazards such as heart-related illnesses and dermatological malignancies.
The primary goal of this framework is to utilize artificial intelligence techniques for computing the Air Quality Index (AQI). To guarantee reliable forecasting, the framework implements robust categorization methodologies. This suggested approach can assist metropolitan regions in controlling and minimizing contamination, providing advantages to both inhabitants and administrative organizations. It predicts the AQI employing parameters such as PM_{10}, $PM_{2.5}$, NO_2, O_3, ***CO, and*** SO_2. ***Functioning as a live monitoring application, it aims to assist individuals in minimizing their contact with atmospheric pollutants.***
Keywords: Data Analytics, Atmospheric Quality, Artificial Intelligence Techniques, Bayesian Classification, K-Nearest Neighbor Algorithm, Data Examination, Guided Learning Models.

I. INTRODUCTION

A fundamental obstacle for intelligent urban environments involves controlling atmospheric contamination, a subject frequently featured in media reports and monitored by meteorological organizations. Although numerous investigations exist regarding pollution management, there exists a significant research void particularly targeting Delhi's atmospheric conditions. This initiative aims to bridge that gap by employing a Machine Learning (ML) framework to forecast air quality. The Significance of ML and AI Machine Learning and Artificial Intelligence (AI) have expanded rapidly and now play essential roles across numerous critical sectors, spanning data extraction and visual recognition to specialized advisory systems. Both emerging startups and well-established corporations have adopted these innovations. The Naïve Bayes (NB) algorithm represents a highly efficient ML technique due to its capability to deliver exceptionally precise outcomes despite minimal training datasets, attributed to its rapid adaptation capabilities. In contrast, Artificial Neural Networks (ANNs) demonstrate greater complexity and prove more suitable for processing extensive data collections, positioning them as an optimal selection for Air Quality Index (AQI) prediction. ML Methodologies for AQI Prediction Numerous diverse algorithms and frameworks have been implemented for this objective, encompassing Neural Networks, fuzzy methodologies, Support Vector Machines (SVMs), SVM regression analysis, fuzzy reasoning, Decision Trees, and K-Nearest Neighbor (KNN) approaches. This specific investigation will employ Neural Networks and Support Vector Machines for AQI forecasting.

II. LITERATURE SURVEY

1) "A Machine Learning Framework for Atmospheric Quality Forecasting in Intelligent Urban Centers"

Authors: Usha Mahalingam, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha, and Giriprasad Kedam Publication Year: 2019 Abstract: This 2019 investigation addresses the expanding difficulty of tracking atmospheric conditions in metropolitan regions. The researchers emphasize that atmospheric quality serves as a crucial measure of contamination and efficient urban governance, since pollutants lead to substantial health complications and ecological damage. The study sought to forecast the Air Quality Index (AQI) utilizing two artificial intelligence methodologies to help prevent contamination from reaching hazardous thresholds. Methodology: The research utilized Neural Networks and Support Vector Machines for its forecasting. Limitations:

- The framework represents a fundamental machine learning instrument and lacks real-time operational capabilities.
- It was constructed to operate exclusively with fixed, pre-collected data.
- The forecasts relied on a limited dataset.
- The overall prediction precision was considerably poor.

2) "Forecasting of Atmospheric Contaminants Using Guided Machine Learning"

Authors: SriramKrishna Yarragunta, Mohammed Abdul Nabi, Jeyanthi P., and Revathy S. Publication Year: 2021 Abstract: This publication emphasizes the severe concern of atmospheric contamination in heavily populated metropolitan regions, recognizing human activities such as transit and power production as primary factors. The researchers observe the growing adoption of machine learning, enabling AI frameworks to examine sensor information and produce dependable predictions. The investigation specifically employs a guided machine learning methodology to forecast the Air Quality Index (AQI), with possibilities for future enhancements to improve forecasting precision. Methodology: The primary technique of the investigation is a Supervised Machine Learning Approach (SMLA). Limitations:

- The investigation was restricted to framework creation and was not deployed as an operational, active system.
- The framework is inappropriate for real-time applications.
- The precision of the framework's predictions was insufficient.

3) "Forecasting of Atmospheric Contamination utilizing Artificial Intelligence Algorithms"

Authors: Jayant Kumar Singh, Amit Kumar Goel Publication Year: 2021 Abstract: This investigation tackles the substantial obstacle of preserving excellent atmospheric conditions in developing, expanding regions where demographic and economic expansion contribute to environmental challenges. The publication highlights the immediate effect of atmospheric contamination—encompassing contaminants such as Nitrogen Oxide, Carbon Monoxide, and particulate substances—on human wellness. The research employs machine learning approaches to predict atmospheric quality metrics. Methodology: The researchers utilized Linear Regression, combined with Python and Jupyter Notebook, to examine SO_2 information. Limitations:

- The framework was developed for static datasets and cannot handle live data streams.
- It is inappropriate for active applications.
- The investigation was constrained by the limited quantity of accessible data.

4) "Intelligent Urban Center Atmospheric Quality Forecasting utilizing Machine Learning"

Authors: Rishanti Murugan, Naveen Palanichamy Publication Year: 2021 Abstract: This publication examines the substantial rise in worldwide atmospheric contamination in intelligent urban centers, with particular emphasis on $PM_{2.5}$—a category of microscopic particulate substance that can trigger severe health complications with extended contact. Referencing traffic bottlenecks as a primary factor in cities such as Kuala Lumpur, the researchers reference the extensive adoption of machine learning algorithms for atmospheric quality prediction and management. The investigation's purpose was to assess the effectiveness of different algorithms for this objective. Methodology: The study evaluated the Multi-Layer Perceptron and Random Forest algorithms. Limitations:

- The framework examined only a limited number of environmental variables.
- It exhibited restricted effectiveness.
- It is inappropriate for real-time surveillance.

III. METHODOLOGY

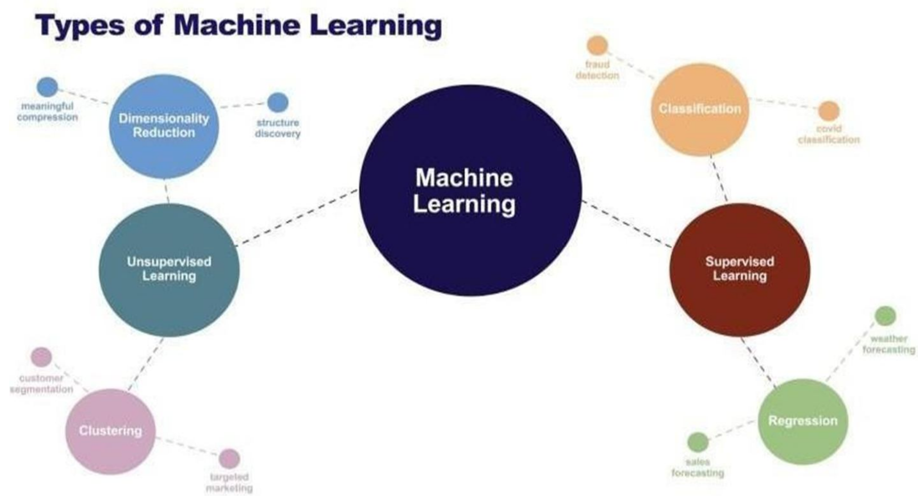
A. *Machine Learning*

Machine learning represents a powerful domain within data science that concentrates on developing frameworks capable of examining and understanding information. Through specialized computational methods, these frameworks can identify trends and correlatio ns, enabling them to generate forecasts or make choices without explicit programming for each situation. It constitutes a methodo logy for addressing intricate challenges that relies heavily on data analysis.

B. *Guided Learning Methodology*

Guided learning represents a predictive modeling technique employed with information collections that include both input characteristics and a related target variable. The objective is to develop a framework that can forecast the target based on the input characteristics. In this learning category, the algorithm gains knowledge from a categorized dataset, where each data instance possesses a predetermined classification. Numerous algorithms exist for guided learning, including K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees, and Support Vector Machines (SVM). The optimal algorithm selection depends on elements such as information characteristics, accessible classifications, and forecasting objectives. The selected algorithm subsequently constructs a framework that can precisely predict results for fresh, previously unseen information, even with some degree of uncertainty.

For this investigation, the Bayesian Classifier and K- Nearest Neighbors (KNN) were selected due to their performance and dependability across different information types and volumes.



C. System Specifications

Persistent industrial growth creates continuous difficulties for tracking and assessing atmospheric quality in metropolitan regions. Evaluating the atmospheric conditions within a particular area is essential for establishing its contamination levels. Atmospheric pollution continues to be a major worldwide issue affecting both the public and regulatory bodies. It results in considerable ecological harm, such as acidic precipitation and climate change, while posing severe health hazards including heart-related illnesses and dermatological malignancies. The primary purpose of this framework is to utilize artificial intelligence techniques for computing the Air Quality Index (A QI). To guarantee precise prediction capabilities, the framework implements effective categorization methodologies. This suggested framework can assist metropolitan regions in controlling and minimizing contamination, providing advantages to both inhabitan ts and administrative organizations. It predicts the AQI utilizing parameters including PM_{10}, $PM_{2.5}$, NO_2, O_3, CO, and SO_2. Functioning as a live monitoring instrument, the framework serves as a beneficial application created to assist individuals in reducing their contact with atmospheric pollutants.

D. Parameters

Values	AQI(Range)	PM10	PM2.5	NO2	O3	CO	SO2	NH3	Pb
1	Good (0-5)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
2	Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5-1.0
3	Moderate (101-200)	101-250	61-90	81-180	101-168	2.1-10	81-380	401-800	1.1-2.0
4	Poor(201-300)	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
5	Very poor(301-400)	351-430	121-250	281-400	209-748	17-34	801-1600	1200-1800	3.1-3.5
6	Severe (401-500)	430+	250+	400+	748+	34+	1600+	1800+	3.5+

Fig 2: Parameters List

IV. ALGORITHMS USED

A. Naïve Bayes Algorithm

The Naïve Bayes algorithm represents a probabilistic categorization technique based on Bayes' principle. Its distinguishing feature is the "naïve" presumption that all attributes within a dataset remain mutually independent. Despite this oversimplification, it frequently demonstrates remarkable effectiveness, especially for applications such as document categorization and unwanted message filtering.

Operating Mechanism

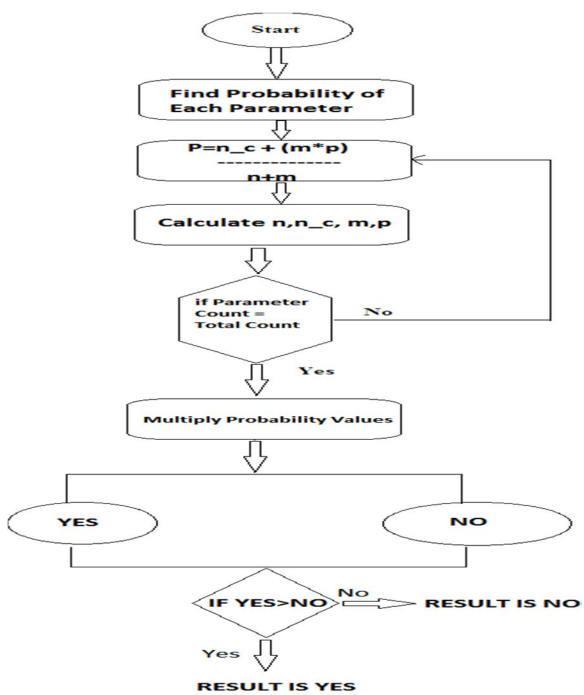
1) Data Acquisition: The procedure commences by obtaining a dataset from sources like files, databases, or remote storage systems.
2) Attribute Probability Computation: The algorithm determines the likelihood of each feature value appearing for all potential classification labels.
3) Formula Implementation: Probabilities are determined using a designated formula, typically incorporating a smoothing method to prevent zero probability outcomes: $P(a_i \mid v_j) = (n_c + m \cdot p) / (n + m)$ Where n represents the total training instances, n_c indicates the count of instances where a particular attribute value and classification value co-occur, p denotes a prior probability estimation, and m represents the equivalent sample volume.
4) Classification Probability Determination: The individual attribute probabilities are subsequently combined with the prior probability of each classification to establish the overall likelihood of a new data instance belonging to that category.
5) Classification Completion: The new data instance receives assignment to the category with the highest computed probability.

K-Nearest Neighbors (KNN) K-Nearest Neighbors (KNN) constitutes a straightforward, non-parametric algorithm employed for both categorization and regression tasks. The core concept assumes that comparable data instances tend to belong to identical categories. It operates by locating the k closest data instances to a new, unclassified instance and subsequently assigning a category based on the predominant classification among those neighbors.

Operating Mechanism

1) Select k: Initially, you must determine the quantity of neighbors to examine, denoted by the parameter k.
2) Distance Calculation: The algorithm computes the distance between the new data instance and all existing data instances in the training collection. Standard measurements for this include Euclidean and Manhattan distances.
3) Locate Nearest Neighbors: It subsequently identifies the k data instances with the smallest computed distances.
4) Determine Classification Labels: The algorithm analyzes the category labels of these k closest neighbors.
5) Category Assignment: The new data instance then receives classification through a majority decision, selecting the most common category among its k nearest neighbors.

B. Navie Bayes Algorithm Flow



FLOW OF NAIVE BAYES ALGORITHM

Fig: 5 – NB Algorithm Flow Diagram

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

C. Naïve Bayes Algorithm Pseudo-code

Input:

Training dataset T,

F= (f_1, f_2, f_3,.., f_n) // value of the predictor variable in testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

   Until the probability of all predictor variables (f_1, f_2, f_3,.., f_n) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

Pseudocode of naive bayes algorithm

Fig: 6 – NB Algorithm Pseudo-code

V. EXPERIMENT RESULTS

A. Naïve Bayes Algorithm

Below are alternative approaches to reformulate the project summary, each emphasizing different aspects.

Variant 1 (Emphasis on Transparency and Effectiveness) For an initiative focused on developing a live monitoring application for community welfare, we utilized Microsoft frameworks to construct a framework. The foundation of this framework consists of a specialized Naïve Bayes library, which we developed using a continuously updating atmospheric contamination dataset. The algorithm's results were exceptional, delivering a 94% correctness rating. It also offered remarkably swift forecasts, requiring only 1,606 milliseconds on average. Our algorithm exhibited strong categorization abilities, reaching 94% exactness and a minimal error rate of 6% for inaccurate classifications.

Variant 2 (Emphasis on Technical Specifications and Performance) This initiative encompassed the creation of a live monitoring application for community benefit, constructed using the Microsoft technology platform. We designed a customized Naïve Bayes library to efficiently handle continuously changing datasets. The developed algorithm showed strong performance indicators, including 94% correctness. Forecasting response time was negligible, with an average duration of 1,606 milliseconds. The algorithm's superior exactness of 94% and an error rate of 6% validate its capability in producing accurate categorizations while reducing mistakes.

Variant 3 (More Streamlined and Business-Focused) We created a live monitoring application for community benefit utilizing Microsoft technologies. The core of the framework is a customized Naïve Bayes methodology, integrated within its dedicated library to process continuously changing atmospheric contamination information. The algorithm showed outstanding results, reaching 94% correctness and a swift average forecasting duration of 1,606 milliseconds. Its categorization strength was validated by a 94% exactness and a 6% error rate, demonstrating its capacity to reduce incorrect classifications.

Constraint	NB Algorithm
Accuracy	94 %
Time (milli secs)	1606
Correctly Classified (precision)	94 %
Incorrectly Classified (Recall)	6 %

Table 1: NB Accuracy

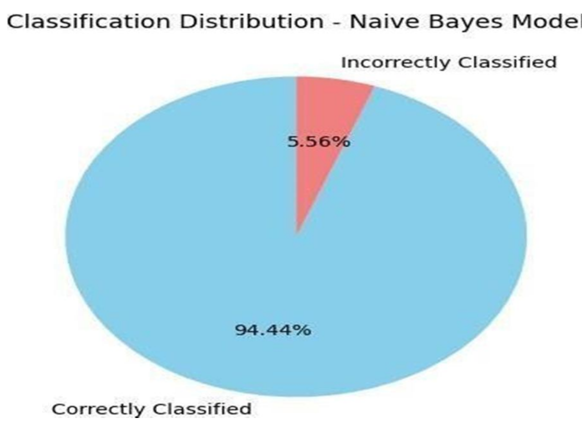


Figure 1: Bar chart showing Accuracy, Correctly Classified, and Incorrectly Classified metrics.

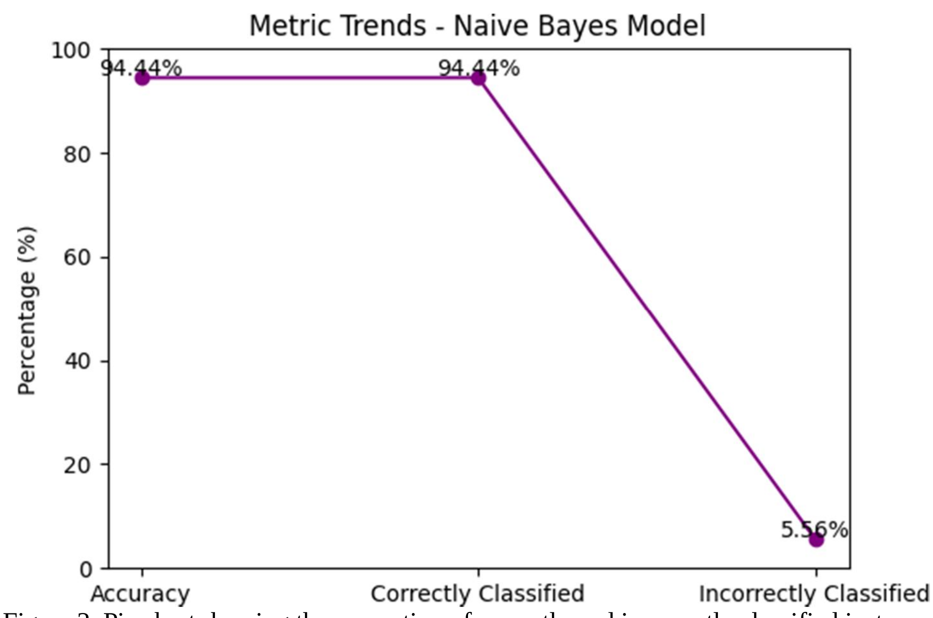


Figure 2: Pie chart showing the proportion of correctly and incorrectly classified instances.

VI. CONCLUSION

Below are several approaches to reformulate the project summary, each highlighting different aspects.

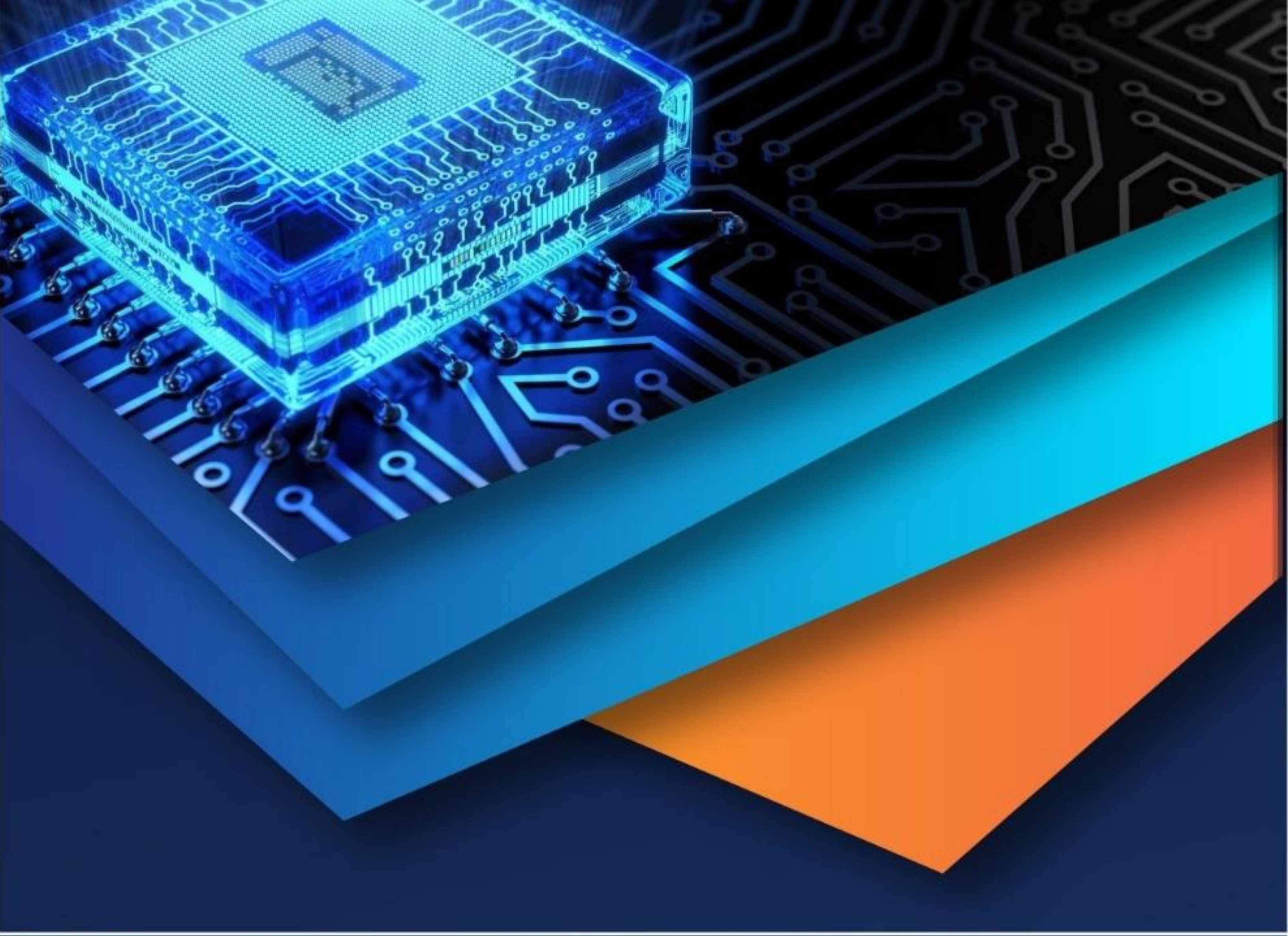
1) Alternative 1 (Emphasis on the Challenge and Resolution) Atmospheric contamination represents a considerable danger to both community wellness and ecological systems, leading to severe complications such as cardiovascular conditions, dermatological malignancies, acidic precipitation, and climate change. To address this challenge, our framework employs an artificial intelligence methodology to forecast the Air Quality Index (AQI). The application delivers live information and practical guidance to assist regulatory authorities in monitoring and minimizing metropolitan contamination.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

2) Alternative 2 (Emphasis on Framework Objective and Operation) To reduce the substantial hazards of atmospheric contamination, encompassing acidic precipitation, climate change, and additional health risks, our initiative created a live application for governmental implementation. This framework utilizes an artificial intelligence algorithm to precisely predict the Air Quality Index (AQI). Through monitoring and controlling contamination levels in urban areas, it delivers essential insights and proposes tactics to safeguard both ecological systems and human wellness.
3) Alternative 3 (More Streamlined and Straightforward) Acknowledging the serious consequences of atmospheric contamination on both wellness and ecological systems, we developed a live artificial intelligence application for governmental organizations. This framework predicts the Air Quality Index (AQI) and provides suggestions to assist in tracking and reducing contamination levels in urban areas, ultimately contributing to controlling environmental harm and enhancing community health.

REFERENCES

[1] Usha Mahalingam et al. (2019): This publication introduces an artificial intelligence framework developed for atmospheric quality forecasting in intelligent urban settings.
[2] SriramKrishna Yarragunta et al. (2021): This investigation examines the prediction of atmospheric contaminants utilizing guided machine learning methodologies.
[3] Jayant Kumar Singh & Amit Kumar Goel (2021): This research describes the implementation of an artificial intelligence algorithm to forecast atmospheric contamination.
[4] Rishanti Murugan & Naveen Palanichamy (2021): This study examines the utilization of machine learning for atmospheric qua lity prediction within an intelligent urban framework.
[5] S. A. Oke (2008): A scholarly review on the topic of artificial intelligence.
[6] O. I. Abiodun et al. (2018): An extensive examination of contemporary implementations of artificial neural networks.
[7] G. Urrútia & X. Bonfill (2010): This publication suggests the PRISMA guidelines as a method to improve the publication of systematic reviews and meta-analyses.
[8] T. Li, M. Hua, & X. Wu (2020): This research introduces a combined framework merging a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network for predicting particulate matter (PM2.5).
[9] H. Liu, Q. Li, D. Yu, & Y. Gu (2019): This investigation concentrates on forecasting the Air Quality Index and atmospheric pollutant levels utilizing diverse machine learning algorithms.
[10] G. Yang, H. Lee, & G. Lee (2020): This publication suggests a combined deep learning framework for predicting particulate matter concentration levels in Seoul, South Korea.
[11] B. Wang et al. (2019): This study presents a framework for atmospheric quality prediction based on a gated recurrent long short- term memory algorithm for implementation within the Internet of Things.
[12] G. Lu et al. (2021): This publication introduces and assesses a novel combined machine learning approach (ORELM-AR) for predicting PM2.5 concentrations.
[13] Wikipedia: A citation to the Wikipedia entry describing the Random Forest methodology.
[14] Vidya (2020): A publication by Analytics Vidya on the subject of artificial intelligence.
[15] Analytics Vidya (2020): A digital resource that describes the confusion matrix in machine learning.
[16] Department of Statistics, Malaysia (2019): A statistical document on the demographic rate in Malaysia.
[17] World of Buzz (2017): A digital publication reporting that inhabitants of Kuala Lumpur experience significant amounts of time in vehicular congestion daily.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

SJIFactor
TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY