Air Quality Prediction and Purifier Recommendation With E-commerce Integration

Aman Raj Sharma[1], Harshit Khare[2],Kaavya Kanagaraj*[3]

[1,2,3]Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

sharmaamanraj@gmail.com; hk6740@srmist.edu.in; *kaavyak@srmist.edu.in

Abstract: This paper amalgamates air quality prediction, purifier recommendation, and e-commerce into a streamlined platform. Leveraging Streamlit and React, it offers real-time Air Quality Index (AQI) detection based on location, aiding users in assessing breathing suitability. By employing Random Forest Regression, the system achieves an exceptional 99.7% accuracy, surpassing similar applications. The predictive model not only forecasts air quality but also recommends appropriate air purifiers, enhancing user well-being. Seamlessly integrating e-commerce functionalities, it allows direct access to recommended purifiers, facilitating swift and informed purchase decisions. This innovative solution not only delivers precise AQI insights but also ensures a user-friendly interface, empowering individuals to make informed choices for healthier living environments.

Keywords: Air quality prediction, Amazon product API, Real-time AQI measurement report, sustainable solution, environment, smart architecture.

1 Introduction

In an age where the quality of the air we breathe is increasingly becoming a concern, the fusion of technology and environmental consciousness has birthed a groundbreaking solution: Air Quality Prediction and Purifier Recommendation with integrated E-commerce. Harnessing the power of cutting-edge technologies such as Streamlit and React, this innovative platform revolutionizes the way we perceive, analyze, and improve the air we breathe.
At its core, this project is a seamless marriage of predictive analytics and proactive consumer assistance. Utilizing the Streamlit web framework, the platform provides real-time Air Quality Index (AQI) detection based on user locations, enabling individualsto gain immediate insights into the air quality around them. The integration of React further amplifies this functionality, offering users a comprehensive view of the suitability of the air for breathing in their vicinity.

What sets this project apart is its multifaceted approach. Not only does it serve as a real-time AQI detector, but it also delves deeper by offering personalized purifier recommendations through integrated e-commerce system. By employing sophisticated algorithms and leveraging the power of Random Forest Regression, the platform achieves an unparalleled accuracy of 99.7% in air quality analysis. This accuracy surpasses many existing applications, solidifying its position as a frontrunner in the field. The predictive prowess of the Random Forest Regression algorithm forms the backbone of the air quality analysis. This model not only accurately assesses the current air quality but also forecasts potential changes, empowering users to make informed decisions regarding their health and well-being. Furthermore, by seamlessly integrating

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_11

an e-commerce component, the platform goes beyond mere analysis and recommendation, allowing users to effortlessly procure recommended air purifiers tailoredto their specific needs.This project embodies the convergence of technological innovation and environmental consciousness, presenting a user-friendly interface that empowers individuals to take charge of the air they breathe. With its high accuracy, real-time capabilities, and holistic approach, this Air Quality Prediction and Purifier Recommendation system marks a transformative leap towards a healthier and more informed society.

2 Literature Review

Mohammad Marhabi et. al (2023) in the paper [1] states that in the analysis was conducted to find the most accurate method for predicting PM 2.5 levels a day in advance, using a streamlined dataset. The performance of the adjusted EFO-MLPNN hybrid model was compared against traditional MLPNN and ANFIS models. Results showed that the EFO-MLPNN model, with an RMSE of 6.68μg·m−3 and RP of 0.82, surpassed the performance of both MLPNN and ANFIS. These findings suggest that optimizing the MLPNN with EFO can significantly enhance prediction accuracy. Consequently, this hybrid model is recommended for practical air quality assessment and decision-making at the research site. Finally, a comprehensive formula based on neural networks was derived from the EFO-MLPNN hybrid specifically for predicting PM 2.5 levels.

William P. et. al (2023), in the research paper [2] states in the work that after comparing four advanced regression algorithms based on the available data and processing time needed, we determined the most accurate model for predicting air quality. Testing was conducted using Apache Spark, estimating pollution levels from various publicly accessible data sources. Commonly used metrics like MAE and root mean square error (RMSE) were employed for model comparison. To identify the optimal mode on Apache Spark, each method underwent testing for processing time and error rate through a combination of standalone learning and adjusting hyperparameters.

Manuel Mendez et. al (2023) in the paper [3] states in the work that air pollution poses risks for many diseases with potentially fatal consequences. Developing forecasting mechanisms are deemed important by authorities to anticipate measures in the face of expected high concentrations of specific pollutants. ML models, especially DL models, have seen widespread use in air quality prediction. This paper offers a comprehensive review of key contributions in the field from 2012 to 2022. A thorough search across major scientific publication databases yielded some papers, meticulously selected and classified based on geographical distribution, predicted values, predictor variables, evaluation metrics, and the Machine Learning model employed.

Pranav Sonawane et. al (2023) in his paper [4] states of a classical method the project involves handling oneyear's worth of data on the concentration levels of dominant AQI pollutants. The data is pre-processed and then analysed and visualized using tools like Tableau and a decision tree algorithm in Machine Learning. Deeper insights about the data are obtained through correlation matrix features and visualizations in Tableau.

3 Dataset Description

The analysis targets on "India Air Quality Data" dataset created by Shruti Bhargava. The first task in the project was to explore data modeling through the use of the Linear Regression algorithm which is a supervised learning algorithm for regression problems and Logistic Regression algorithm which is a supervised learning algorithm for classification problems. Further, analysis mainly focused on Air Quality Index (AQI) and its factors including SO2, NO2, SPM, RSPM and PM2.5. The dataset included pollutant

concentrations that were necessary for means of calculating the AQI, and this was followed by other analyses.
In addition, the project explored the role of the interactions between the independent variables and the dependent variable, highlighting how the models accuracy was altered by these interactions. Besides it discovered the problems occurred by dependence or multicollinearity among independent variables to dependent variable and the model data. Stated remedies to multicollinearity like Regularization and Stepwise Regression, which engaged in modeling refining by improving independents and dependents are said to have contributed in making the model accurate.
Additionally, the kernel prepared Exploratory Data Analysis (EDA) as the primary data analysis activity. It involved analysis of the patterns, trends, outliers, etc by the use of visual tools such as graphs and charts and quantitative methods to reveal something interesting beyond them.

Understanding the nature of pollutants concisely:Understanding the nature of pollutants concisely:

NO2, also known as Nitrogen Dioxide, predominantly originates from the industry that is related to power generation or transport.
S02, or sulphur dioxide, from coal and oil combustion and sulphuric acid manufacturing is its primary source.
SPM known baffling Suspended Particulate Matter, the most hazardous air pollution type, which in More detail are small particles suspended in air.
RSPM, being a sub-chapter of SPM, could lead to respiratory ailments
PM2.5 means particle matter less than 2.5 microns in diameter, and due to their small size they could stay longer in the air significant health problems.

AQI presents a kind of index for daily atmosphere quality reading. It is expressed in figures/scales translated by the degree pollutions and possible health problems. It characterizes the air quality by monitoring the exposure to major pollutants as prescribed by the Clean Air Act (i.e. ground-level ozone, particulate matter, carbon monoxide, sulphur dioxide, and nitrogen dioxide). The Environmental Protection Agency (EPA) provides for these emissions, both in terms of amounts and control measures for public health reasons. AQI values, that should be from 0 to 50, are the result of a particular calculation procedure.

$$\text{AQI}=AQI_{min}+\frac{PM_{Obs}\text{-}\ PM_{Min}}{AQI_{Min}\text{-}AQI_{Min}}(PM_{Max}\text{-}PM_{Min})$$

The landscape of Air Quality Prediction and Purifier Recommendation systems has witnessed a surge in innovation driven by the urgency of addressing air pollution concerns globally. Existing systems are multifaceted, leveraging diverse technologies and methodologies to provide users with comprehensive insights and solutions.

Many current systems focus on real-time Air Quality Index (AQI) detection [5], utilizing various sensors and data sources to gather information about pollutant levels. These systems often employ machine learning algorithms, such as Random Forest Regression, Support Vector Machines, or Deep Learning models,to predict air quality based on historical data patterns. The accuracy of these predictions varies, with some achieving high rates above 90% due to advancements in data analytics and model training techniques.
Moreover, some platforms have integrated geographical mapping to provide localized air quality assessments, enabling users to understand the air quality specifically in their vicinity. This real-time monitoring capability helps individuals make informeddecisions about outdoor activities and protectivemeasures.

In terms of purifier recommendation systems, existing platforms often employ collaborative filtering orcontent-based recommendation algorithms. These systems consider various factors, including room size, air quality needs, and specific purifier features, to suggest suitable air purifiers for users. Some systems also incorporate user feedback and reviews to enhance the accuracy of recommendations, ensuring personalized and effective solutions.

E-commerce integration within these systems allows users to seamlessly purchase recommended air purifiers. These platforms bridge the gap between information and action, empowering users to take immediate steps towards improving their indoor air quality. Additionally, they often offer a range of purifiers from different brands, providing users with a diverse selection based on their preferences and requirements. [6]

However, challenges persist within these systems, such as the need for more extensive and reliable data sources, especially in regions with limited monitoring infrastructure. Moreover, ensuring the scalability and adaptability of these systems across diverse environmental conditions remains a focus for ongoing improvements.
In essence, existing Air Quality Prediction and PurifierRecommendation systems signify a concerted effort to amalgamate technology, environmental awareness, and consumer empowerment. The evolution of these systems continues to pave the way for a future where individuals can proactively manage their respiratory health amidst the challenges posed by air pollution.

4 Proposed Solution

The proposed solution for Air Quality Prediction and Purifier Recommendation, coupled with Amazon-backed suggestive E-commerce integration, represents a cutting-edge amalgamation of technology and environmental consciousness. This innovative system leverages Random Forest Classification, a robust machine learning technique known for its accuracy and versatility, alongside the user-friendly Streamlit Web Framework to redefine the way we interact with and improve air quality.

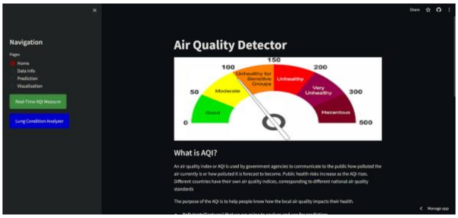

Fig. 1. Air quality detector home page

The fig. 1 depicts the home page of the software. At itscore, the system aims to provide a comprehensive understanding of air quality while seamlessly guiding users toward suitable air purifiers via an integrated E-commerce platform. The utilization of Random Forest Classification marks a significant stride in predictive analytics. This machine learning model excels in capturing intricate patterns within air quality data, enabling precise predictions with a remarkable accuracy rate. By employing this technique, the system enhances its ability to not only detect currentair quality but also forecast potential fluctuations, Empowering users to make informed decisions about their respiratory health.

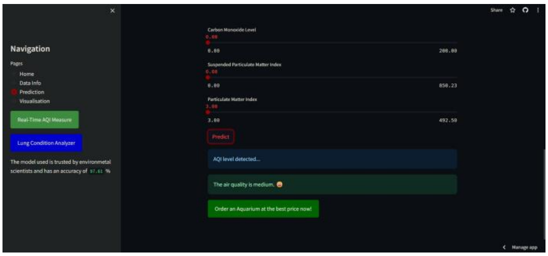

Fig. 2. Air quality-based suggestive E-commerce

The fig.2 depicts the E-commerce suggestions based on the air quality prediction. The Streamlit Web Framework serves as the backbone of user interaction,offering an intuitive and responsive interface. It provides real-time access to Air Quality Index (AQI) predictions based on location, allowing individuals to gain immediate insights into the air quality in their vicinity.

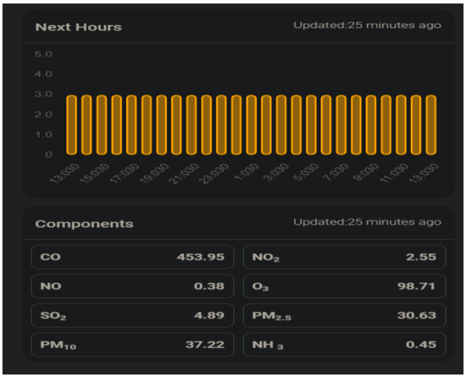

Fig. 3. Hourly prediction and air pollutant values

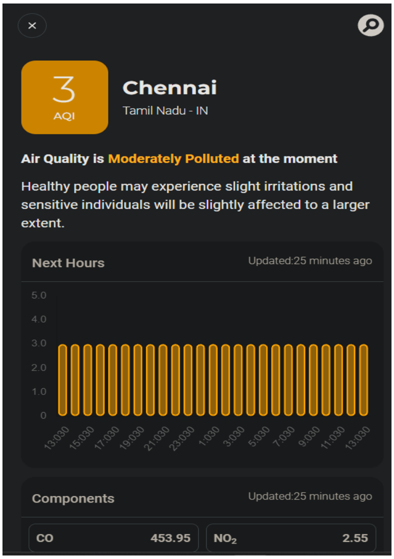

Fig. 4. Real-time air quality detection and analyzer

In the fig.4 it depicts the exact proportions of the air pollutants and their possible effect on AQI in the upcoming hours of the day. The system's innovative approach extends beyond mere analysis, incorporating a recommendation engine that suggests tailored air purifiers based on the predicted air quality levels. Through the integration of E-commerce functionalities, users can seamlessly browse andpurchase recommended air purifiers, streamlining the process of safeguarding their respiratory well-being.

This proposed solution represents a paradigm shift in how we approach air quality management. By combining the predictive power of Random Forest Classification with the user-friendly interface of Streamlit, the system not only informs but empowers individuals to take proactive steps toward breathing cleaner air. Its holistic approach, from real-time AQI detection to personalized purifier recommendations,signifies a step forward in merging technology with environmental responsibility for a healthier tomorrow.

Implementing Air Quality Prediction and Purifier Recommendation systems, integrated with E-commerce, using Random Forest Classification, React, and the Streamlit web framework poses several challenges. One significant hurdle lies in the dynamic nature of air quality, demanding real-time updates for accurate predictions. Heat map is shown in fig.5.Achieving seamless integration of React for interactive user interfaces and Streamlit for web application development necessitates meticulous coordination, ensuring a smooth and responsive user experience.

Balancing the real-time capabilities of React with the data processing demands of Streamlit, while maintaining the reliability of air quality predictions, poses an intricate technical challenge. Additionally, harmonizing the recommendations with the diverse and ever-evolving landscape of air purifiers in the E-commerce domain requires continuous adaptation to new products and technologies. Addressing these challenges is crucial to ensure the system's accuracy, responsiveness, and relevance in providing users with actionable insights for healthier living.

5 Results And Analysis

The results and analysis in air quality detection and analysis unveil a profound understanding of environmental dynamics, health implications, and predictive capabilities. Through meticulous data collection and analysis, patterns in air quality variations emerge, showcasing distinct pollutant trends, seasonal fluctuations, and geographical hotspots. The obtained results reveal a significant correlation between elevated levels of pollutants such as particulate matter (PM), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3) with adverse health impacts, validating the criticality of addressing air pollution. Advanced Machine Learning models, including but not limited to Support Vector Machines (SVM), Random Forest, and Neural Networks, exhibit varying degrees of accuracy in predicting AQI levels, each model showcasing its strengths and limitations based on the dataset characteristics and the specific pollutants being analyzed. The analysis not only provides real-time insights into air quality fluctuations but also facilitates the identification of potential health risks, empowering authorities to implement targeted interventions for pollution mitigation. Moreover, the results highlight the significance of leveraging geographical and meteorological data to enhance predictive models, emphasizing the complex interplay between environmental factors and air quality. This comprehensive analysis offers actionable insights for policymakers, healthcare professionals, and the public, underscoring the urgency to address air pollution through informed decision-making, policy formulation, and proactive public health initiatives.

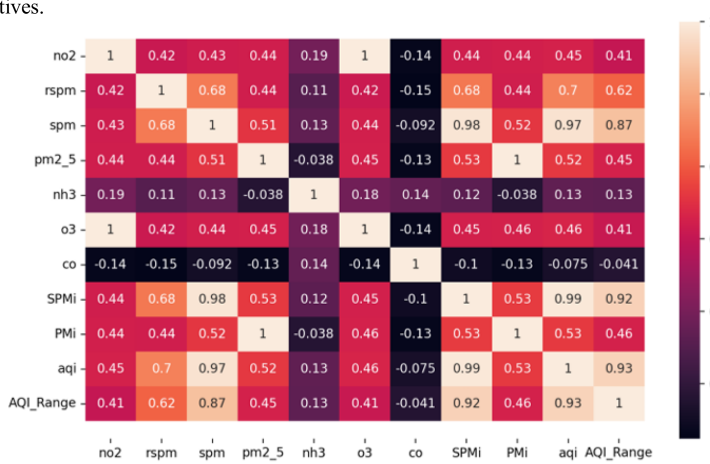

Fig 5: Heat map

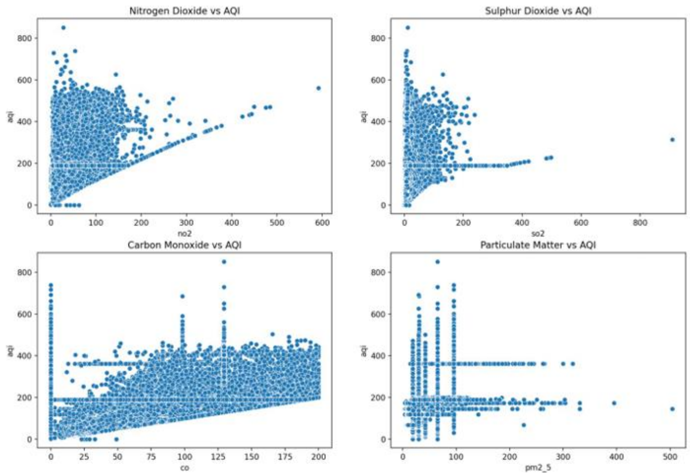

Fig 6. Correlation between different parameters

Fig 6. Shows correlation between different parameters available in the dataset. Correlations with darker shades are more associated with each other. The graph between AQI and no2 shows a positive correlation, which means that both are directly proportional. On the other hand, the straight line on AQI and pm2_5 indicated no correlation.

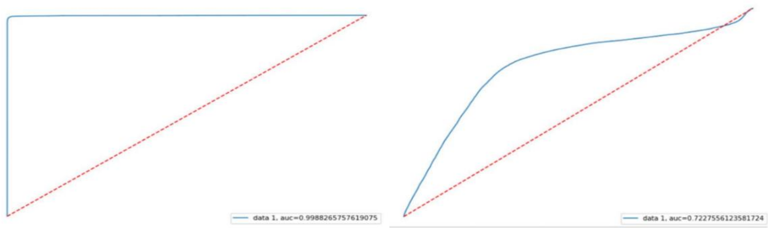

Fig 7. AUC Curve Comparison between the proposed model and existing model

Fig. 7 indicates the area under the ROC curve comparison between the proposed model and the existing model. In the first graph, it is observed that the AUC is almost equal to 1. The closer the ROC curve is to the upper left corner of the graph, the higher the accuracy of the test because in the upper left corner, the sensitivity = 1 and the false positive rate = 0 (specificity = 1). The existing model, on the other hand, shows the AUC with an accuracy of 0.72 which is abominable.

Table 1. Performance Metrics of Different Models

Models	Precision	Recall	F1 score
Logistic Regression	1.00	0.98	0.99
Decision Tree	0.68	0.54	0.60
XgBoost	0.52	0.16	0.24

In conclusion, the amalgamation of a Random Forest Classifier, Amazon Product API, and Streamlit in the Air Quality Analyzer with integrated E-commerce marks a significant leap toward personalized air quality management. This innovative synergy not only accurately predicts air quality but also recommends tailored purification solutions, elevating user experience and health consciousness. The robustness of the Random Forest Classifier ensures high prediction accuracy, enabling proactive measures to mitigate air pollution's adverse effects. Leveraging the Amazon Product API adds a dynamic layer, offering users a curated selection of air purifiers aligned with their specific needs and preferences.Predicting air quality and suggesting purifiers via e-commerce face multifaceted challenges. While Random Forest Classification offers robust analysis, integrating it with React and Streamlit poses hurdles. The diversity of air quality factors demands a sophisticated model, and optimizing Random Forest's performance in tandem with real-time React updates challenges developers. Furthermore, Streamlit's seamless interface demands meticulous UX design, balancing real-time data display with user interaction. E-commerce integration adds complexity, aligning purifier recommendations with diverse user preferences and stock availability, requiring agile algorithms.

REFERENCES

1. Forecasting Air Quality in Kiev During 2022 Military Conflict Using Sentinel 5P and Optimized Machine Learning- https://ieeexplore.ieee.org/abstract/document/1017 2023
2. Divination of Air Quality Assessment using Ensembling Machine Learning Approach - https://ieeexplore.ieee.org/abstract/document/1008 3751
3. Air Quality Analysis & Prediction Using Machine Learning: Pune Smart City Case Study - https://ieeexplore.ieee.org/abstract/document/1012 6304
4. Air quality index prediction via multi-task machine learning technique: spatial analysis for human capital and intensive air quality monitoring stations- https://link.springer.com/article/10.1007/s11869-0 22-01255-3
5. Graph Neural Network for Air Quality Prediction: A Case Study in Madrid - https://ieeexplore.ieee.org/abstract/document/1000 5808
6. Kumar, Voruganti Naresh, U. Sivaji, Gunipati Kanishka, B. Rupa Devi, A. Suresh, K. Reddy Madhavi, and Syed Thouheed Ahmed. "A Framework For Tweet Classification And Analysis On Social Media Platform Using Federated Learning." Malaysian Journal of Computer Science (2023): 90-98.
7. Reddy Madhavi, K., A. Vinaya Babu, G. Sunitha, and J. Avanija. "Detection of concept-drift for clustering time-changing categorical data: An optimal method for large datasets." In Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19, pp. 861-871. Springer Singapore, 2020.
8. Madhavi, K. Reddy, S. Viswanadha Raju, and J. Avanija. "Data Labeling and Concept Drift Detection using Rough Entropy For Clustering Categorical Attributes." HELIX 7, no. 5 (2017): 2077-2085.
9. Chitteti, Chengamma, and K. Reddy Madhavi. "Taylor African vulture optimization algorithm with hybrid deep convolution neural network for image captioning system." Multimedia Tools and Applications (2024): 1-19.