International Journal of Medical Toxicology & Legal Medicine Volume 27, No. 5, 2024
e-ISSN: 0974-4614
p-ISSN: 0972-0448
https://doi.org/10.47059/ijmtlm/V27I5/027

Predicting Air Quality Index in Real Time and Classifying Its Health Effects: Advancements in Machine Learning

Narendra Babu Pamula[1], Dasari Anusha[2], Dudla Anil Kumar[3], Ruqsar Zaitoon[4], K Jairam[5], M.V.P. Umamaheswara Rao[6]

[1]Sr. Asst.Professor in Department of Artificial Intelligence & Data Science, Lakireddy Bali Reddy College of Engineering(A), A.P., India, Email:naren.pamula@gmail.com
[2]Assistant Professor in Electronic and Instrumentation Engineering, Velagapudi Ramakrishna Siddhartha Engineering College Deemed to be University,A.P, India
[3]Sr. Asst.Professor in Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering(A), A.P., India
[4]Assistant professor,Department of Information Technology,Vardhaman College of Engineering, Hyderabad, Telangana, India
[5]Assistant Professor in Computer Science and Engineering, SRK Institute of Technology, A.P., India
[6]Associate Professor, Department of Computer Science and Engineering, NRI Institute of Technology(A), A.P., India.

Received: 17.09.2024 Revised: 10.10.2024 Accepted: 07.11.2024

ABSTRACT
Accurate and early forecast of air quality is therefore very important since air pollution seriously compromises public health and environmental sustainability. This work includes a thorough investigation on real-time Air Quality Index (AQI) prediction utilizing advanced machine learning approaches and classification of its health impacts. To build strong predictive models, the proposed method combines several environmental information including meteorological factors and pollution concentrations. Urban and rural environments are guaranteed the real-time framework's applicability since it is maximized for scalability and reactivity. The study also assesses the performance of the model by means of large-scale experiments, therefore contrasting interpretability, latency, and accuracy. The results show how well machine learning might improve air quality monitoring systems and support proactive actions to lower health hazards. This study emphasizes how important artificial intelligence is in tackling public health and environmental concerns of great relevance. The quality of the air gets daily declining in recent years. To control and stop regional air pollution, exact forecast of AQI concentration is really useful. It is used to quantify how human health suffers under air pollution. One can find the air quality index in numerous methods. To find the AQI in this work we apply PM2.5, PM10, NO, NO2, NOX, NH3, CO, SO2, O3, benzene, Toluene characteristics. Among the most fascinating approaches to forecast and evaluate AQI are machine learning methods. This work applies XGBoost Regressor, Catboost Regressor, Random Forest Regressor algorithms. One can hone the most effective technique to identify the ideal answer. Thus, the work of this article consists on thorough investigation and application of new technologies such SMOTE to guarantee the best feasible solution to the issue of air quality.

Keywords: Air Pollution Forecasting, Machine Learning, Health Risk Classification, Health Impact Assessment.

INTRODUCTION
For human survival, air is essential. For our own safety, we need to study its characteristics. As a result of air pollution, millions of people experience physiological ailments or even respiratory fatalities. The most serious environmental problem, according to research, is air pollution. The rapid industrialization has led to the release of harmful gasses, which in turn has generated a tremendous increase in the population. The contamination of harmful substances has a devastating effect on our well-being. Since air quality has a significant impact on public health and wellbeing, research into the real-time prediction of the Air Quality Index (AQI) and associated health impacts is crucial. The Air Quality Index (AQI) is a standardized measure for expressing levels of air pollution and the related health concerns. However, the precision and timeliness required for sound decision-making may be missing from traditional methods of AQI surveillance and prediction. Machine learning (ML) is quickly becoming a game-changing answer to these issues. With the use of massive environmental datasets, ML models can quickly and accurately predict AQI levels and categorize health risks associated with various

pollution levels. These advancements permit prophylactic measures to lessen health impacts, particularly for susceptible populations. The air we breathe is in terrible shape due to this unchecked pollution. For the purpose of quantifying and conveying air pollution levels, the AQI is utilized. The air quality forecast aims to detail and assess the current and future conditions of air pollution, environmental quality trends, and the most significant pollutants and pollution sources in terms of their dynamic changes. Sensors and the IoT used to be the primary sources of data and predictions. This can be used as a starting point for suggestions about ways to improve the environment and stop it from getting worse. Because of this, being able to anticipate the AQI index in advance is vital, since it allows people to take precautions and reduce losses. In today's world, when air pollution is becoming more of a problem, reliable weather predictions are more important than ever. Hence, it is critical to provide precise predictions of air quality. Hence, it is essential to be able to forecast the air quality in advance so that people can take precautions in a timely manner to limit their losses. Pollutants in the air typically come from a variety of places, including vehicles, volcanic eruptions, and factories. Also, different regions have different levels of pollution, which means that different things contribute to smoky weather to different degrees. There is a specific association between numerous atmospheric elements, and the atmospheric system as a whole is very complex. The meteorological variables linked to the air quality index (AQI) can be determined by factor analysis. As a result, AQI forecasts are challenging. Oxygen monoxide (O3), sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter (PM)10, particulate matter (PM2.5), ammonia (NH3), benzene, toluene, xylene, nitroxygen (NHX), and nitrogen (NO) are the twelve variables that make up the air quality index (AQI). The Air Quality Index (AQI) is computed in other applications using the following six pollutants: PM10, PM2.5, SO_k, NO_k, CO, O_t, and benzene. Nevertheless, other factors, such as data availability, measuring methods, and monitoring frequency, contribute to the precise selection of contaminants. When the Air Quality Index (AQI) is high, it means that the air is very polluted, which is bad for your health. Air quality can be tracked in real-time with the help of the AQI. Additionally, AQI data was collected hourly and daily by a number of meteorological stations. The suggested work makes use of this data that has been mined and collected. Hence, we drew on a dataset that includes AQI readings for a number of Indian towns. In order to choose the most accurate one, we run three separate regression analyses and compare their results. The proposed study evaluates the dataset's performance both before and after applying the SMOTE algorithm. We used SMOTE, which is the biggest invention. We utilized SMOTE to balance the dataset and investigated its effects, unlike previous works. Future AQI levels can be predicted using the suggested approaches, which can be used as a warning and to highlight the importance of reducing air pollution. This study explores the integration of ML techniques for real-time AQI prediction and health risk classification, focusing on recent breakthroughs, limitations, and future development potential.

LITERATURE SURVEY

The SVR with an LSTM model is used to build the classification mechanism, which predicts the principal location AQI of the metropolitan region using the deep learning mechanism. The Air Quality Index (AQI) informs the public about the state of the air in a specific metropolitan area. The air quality index is measured in urban areas, and the precise reading was acquired by means of deep learning [1]. Five separate models using three separate data sets are examined and evaluated in a separate study. The three databases were interconnected, but they presented data in distinct ways and had varied scopes. Various models are applied to various datasets. In terms of accuracy and root-mean-squared error (RMSE), the PLS model outperformed the KNN model with 10 iterations of tuning length CV. precision as well as area under the curve density [2]. Using the Italian urban pollution concentration forecast and the Beijing AQI forecast as examples, they constructed regression models to anticipate air indicators using machine learning algorithms in another work. According to the findings, RFR-based models outperform SVR-models in terms of quality of results [3]. Machine learning techniques for air quality index prediction in the Chennai region utilizing the MLR model. Outcomes from the MLR model are satisfactory. The ARIMA algorithm is then employed, and it outperforms MLR on the provided dataset. The accuracy that ARIMA provides is 95% [4]. Using the given dataset, they applied the KNN method to forecast the AQI values. There are two varieties of the dataset's algorithm: one for raw data and one for normalized data. After that, they used RNN on the datasets, and the results varied. The accuracy for both the raw and normalized data sets is 92.86% when using KNN [5]. We employ LSTM and ILSTM in this paper. Overall, CNN-ILSTM outperforms other deep learning models and more traditional regression models, such as SVR, RFR, and MLP, as well as LSTM, GRU, ILSTM, CNN-LSTM, and CNN-GRU [6]. Still another method for AQI determination in Nanjing using AR data. The model provides a practical method for predicting values. The hybrid KF-AR model outperforms the individual AR models [7]. By utilizing various assessment measures such as R2 score and RMSE, SVM with RBF kernel outperformed other models and made numerous adjustments to the SVM model in order to forecast the AQI data of Ahmedabad city. Researchers can examine the suggested model for AQI data prediction in the future from a variety of locations, including local or state level analyses. It has the potential to be expanded to forecast other pollution indices at varying levels [8]. Systems [9]. Predicting AQI from parameters and AQI is done using algorithms such as MLR, ANN, DT-ANN, DT, and MLR-ANN. Algorithms differ in their level of precision. When compared to all other models, DT-ANN produces the most accurate results. The accuracy of AQI predictions was enhanced by DT [10]. Analyzing pollution levels in Los Angeles and Houston, the study aimed to construct a model for predicting the Air Quality Index (AQI). Predicting AQI with more precision was possible using a stacked ensemble of separate predictors. The cascaded forward group outperformed the support vector regression (SVR) set when comparing the two fundamental learning sets. A typical problem for neural network approaches that try to converge to locally optimal solutions is properly predicting AQI from highly nonlinear data; nevertheless, the study discovered limits in this area as well. To tackle this, researchers in the future could look at more advanced deep learning methods like restricted Boltzmann machines and long-short-term memory (LSTM) networks, which can model complicated mappings well by layering features. These methods have the potential to enhance air quality management by enhancing AQI predictions through the integration of low-level features with high-level representations [11]. In order to forecast the AQI, a hybrid model known as VMD-SE-LSTM was created. To begin, the original AQI set was decomposed using the VMD technique. Then, to fix the over-decomposition and alleviate the computational burden, SE was used to merge the VMD subset. The following step was to forecast each recombination subsequence using a long short-term memory (LSTM) neural network, which is renowned for its learning capabilities and capacity to remember past data. According to the evaluation criteria, the VMD-SE-LSTM hybrid model achieved near-optimal performance in AQI prediction, surpassing five reference models: BP, LSTM, EEMD-BP, EEMDLSTM, and VMD-SE-BP [12]. By utilizing hourly air quality data, the accuracy of representing urban air quality is enhanced in this work. The first step is to use transfer entropy analysis to choose six meteorological variables that have a major impact on air quality. Incorporating aging forecasting into the mix, the study compares the accuracy of predictions made by LSTM, GRU, and BP neural network models over time horizons ranging from zero to forty-eight hours. When measuring two variables side by side, the RMSE is the statistical measure of choice. The results demonstrate that LSTM outperforms standard neural networks in terms of prediction accuracy and reliability as the prediction horizon rises. Short-term and temporal forecasting are two other areas where LSTM excels [13]. Finding a way to forecast AQI using DTMC—which represents daily variations in air quality—was the primary goal of the project. He found that the most common air pollutants in Taipei were O3, NO2, and PM10. To evaluate the air quality and its health impacts on a daily basis, precise AQI forecasts are essential. Human activities in Taipei are strongly associated with these contaminants, according to the research. Those findings could be utilized by governments to develop efficient strategies for reducing air pollution. Reducing emissions from motor vehicles and air pollution can be achieved by encouraging behaviors like carpooling, taking public transportation, cycling, or walking [14].

DATA DESCRIPTION

The below provided link is been used for this work.

https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india

The dataset includes air quality index (AQI) readings for various cities in India, broken down by day. These numbers cover the years 2015 to 2020. The initial dataset had 16 columns and 29531 rows. The AQI Bucket can return results ranging from Good to Very Poor to Severe, with a moderate possibility of being satisfactory. With the null values removed from Xylene, Excel can pull out a plethora of cities from the original dataset. The cities that make up this cluster are: Amaravathi, Bengaluru, Hyderabad, Mumbai, Ahmedabad, Lucknow, Kolkata, Chennai, Visakhapatnam, Thiruvananthapuram, and Delhi. After the dataset was cleaned, there were a few missing values for attributes including PM2.5, PM10, AQI, and NH3. Those variables were filled using the mean values. Additionally, there are now 18399 rows and 14 columns in the updated sample dataset. Since the original dataset was unbalanced to begin with, the synthetic minority oversampling method (SMOTE) was employed to rectify the situation. As an oversampling technique, SMOTE creates fake samples for underrepresented groups. A skewed class distribution, caused by an imbalanced dataset, can impact the model's accuracy. We used SMOTE to fix the data imbalance problem, namely by oversampling the positive class label, thus our results were more accurate. Using SMOTE, a method based on closest neighbour analysis, we can successfully boost the dataset's representation of examples from minority classes. With the help of SMOTE, we were able to improve the data balance in our dataset, which had six positive classes and twelve negative ones to begin with. This delicate balancing effort improves algorithm efficiency while reducing the risks of overfitting. Using SMOTE was a series of iterative steps: finding feature vectors and their neighbors, determining their difference, adding random variations, and finally, creating synthetic data points along the newly formed line segments. To add variety without duplication, SMOTE distinguishes itself from simple replication by generating synthetic instances that differ marginally from the original data. Moderate, satisfactory, good, poor, extremely poor, and severe were the six AQI Bucket designations that we monitored in our study. We addressed the disparities that were initially evident by observing a large convergence in label frequencies through numerous rounds of the SMOTE algorithm. For example, suitable modifications, represented as 0 values, were made in Bangalore since there were no "severe" label values and in Delhi because there were no "good" label values in the AQI Bucket column. By addressing imbalances in the dataset, our study utilizes SMOTE to improve model accuracy.

Inaccuracies in the model may result from skewed class distributions caused by such imbalances. Improved balanced accuracy and detection rates are signs that SMOTE is helping to generate more accurate models by encouraging data balance. Crucially, SMOTE counters the overfitting tendencies often linked to random oversampling methods by creating synthetic data points that promote diversity without replication. Classifier generalization and performance are both improved by SMOTE's example generation, which creates a more open and flexible decision boundary. Improving the reliability and resilience of our prediction models relies heavily on this skill. Figure 1 displays the results of a comparison study of several cities' imbalanced data values.

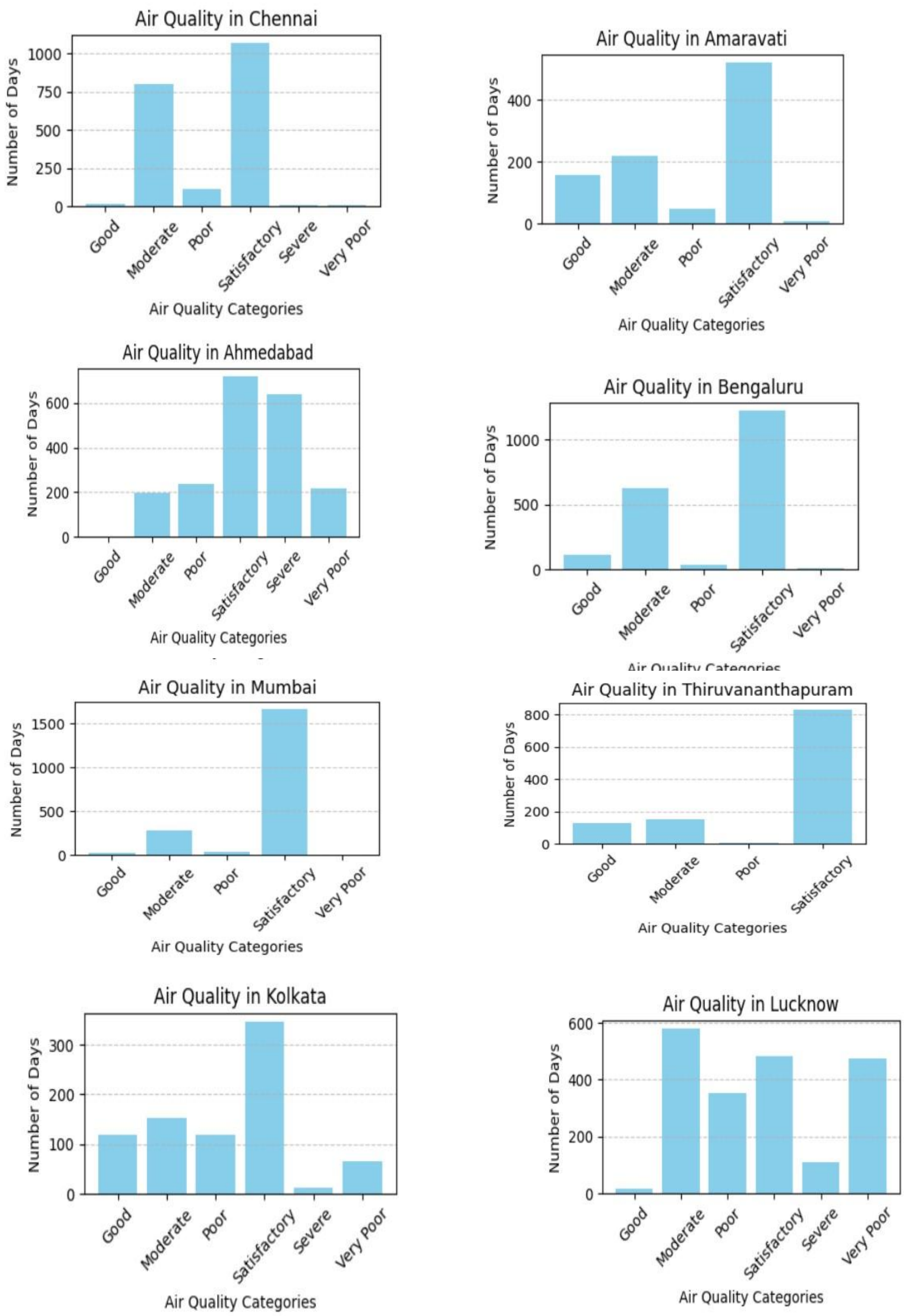

Figure 1: AQI Comparative analysis of different cities with their Imbalanced data values**.**

Because of differences in industrial operations, traffic pollution, weather, and urban planning, the Air Quality Index (AQI) values and related health effects vary substantially throughout Indian cities. Amaravathi, New Delhi, Bangalore, Kolkata, and Hyderabad's air quality index trends and their health effects are reviewed here in

Table 1: we can show the Air Quality Index (AQI) levels and their health impacts.

<table>
<tr><th>City</th><th>AQI Level</th><th>Health Impact</th><td rowspan="6">Concerns for General Health

Prolonged exposure to high AQI levels is associated in all cities with:

Irritation of the eyes, pain in the throat, and headaches are some of the acute health effects.
Long-Term Consequences on Health: Deterioration of lung function, cardiovascular illnesses, and an Increased Risk of Mortality.

Suggestions for Enhancing the Health of Citizens

Wear N95 masks on days when pollution levels are high.
When pollution levels are high, it's best to stay indoors.
Advocate for renewable energy programs and stronger regulations on vehicle emissions.
Make public transportation better in order to lessen traffic jams.
Make more room for green areas so they can filter the air naturally.</td></tr>
<tr><td>New Delhi</td><td>PM2.5 and PM10 levels frequently exceed safe limits.</td><td>• More cases of respiratory illnesses (bronchitis, asthma).
• Cardiovascular disease occurs more frequently.
• Decreased pulmonary development in growing youngsters.
• India has one of the highest annual fatality rates linked to air pollution.</td></tr>
<tr><td>Bangalore</td><td>Bangalore typically keeps its air quality index levels reasonable.</td><td>• The number of people experiencing difficulty breathing increases during rush hour.
• Seasonal changes in air quality can trigger asthma attacks and allergy reactions.
• Reduced lung function in susceptible populations has been associated with long-term exposure.</td></tr>
<tr><td>Kolkata</td><td>Because of the high concentrations of PM10, NO2, and SO2, Kolkata frequently has low AQI.</td><td>• More and more people are suffering with COPD, or chronic obstructive pulmonary disease.
• Cardiovascular and cerebrovascular disorders are prevalent.
• A rise in the number of people admitted to hospitals due to ailments caused by air pollution, especially children and the elderly.</td></tr>
<tr><td>Hyderabad</td><td>There is a wide range of air quality index readings for Hyderabad, from fair to poor.PM2.5 and NO2 are the primary pollutants.</td><td>• Asthma and respiratory allergy cases are on the rise.
• Extended exposure is associated with an increased risk of ischemic heart disease and stroke.
• Proximity to industrial zones puts slum communities at risk.</td></tr>
<tr><td>Amaravathi</td><td>As a developing metropolitan hub, Amaravathi usually has moderate air quality index (AQI).</td><td>• When stubble is burned, it might irritate the respiratory system.
• Compared to highly industrialized cities, there are comparatively fewer long-term health consequences.
• Growing urbanization and vehicle density pose a threat to already high pollution levels.</td></tr>
</table>

METHODOLOGY

It is important to look at the specific demographics, exposure levels, and weather patterns of big cities like New Delhi, Bangalore, Kolkata, and Hyderabad to appreciate the relevance of the air pollution index (AQI). In this piece, we take a look at how the Air Quality Index (AQI) impacts city people' health. Four cities—New Delhi, Bangalore, Kolkata, and Hyderabad—had their air quality index readings compared in this study.The objective is to identify the most accurate and efficient algorithm among the three. The study aims to present the analysis in a concise and effective manner, facilitating the discovery of insightful information. These cities, characterized by high population density, serve as representative samples of pollution levels in major South Asian urban centers. The selection of these cities over others was made to maintain the paper's focus and prevent unnecessary lengthening. Thus, by focusing on India's major urban centers, this research provides valuable insights into pollution levels across different regions of the country. Here we provide the comparative analysis of different cities.

Synthetic Minority Oversampling Technique (SMOTE) Algorithm:

By creating synthetic samples for the minority class, SMOTE is an oversampling strategy that aims to correct imbalanced datasets. An overfitting problem often seen with arbitrary oversampling approaches can be reduced with the use of SMOTE's dataset balancing techniques.

Random Forest Regression (RFR) Algorithm

When it comes to classification and regression, RFR is one of the most used supervised machine learning techniques. It builds decision trees from various samples using class voting for classification and averaging for regression.

CatBoost Regression (CR) Algorithm:

Yandex has developed CatBoost Regression (CR), which provides a free and open-source gradient boosting framework. To improve model performance and interpretability, CatBoost takes a different approach than typical approaches by adopting a permutation-based alternative to handle categorical features.

XGBoost Regressor (XGBR) Algorithm

Extreme Gradient Boosting (XGBoost), a state-of-the-art supervised learning method, is very good at regression and classification. The prediction model is refined through iterative application of gradient boosting, which minimizes a predefined loss function. Iteratively fixing the faults made by preceding learners using the XGBoost algorithm is how a cluster of unsuccessful learners, typically decision trees, is built. The XGBoost package is an excellent choice for effective and scalable administration of huge datasets containing millions of instances and features. Overfitting can be mitigated via regularization techniques, particularly L1 and L2 regularization.Now that XGBoost can automatically deal with missing data and categorical features, it's much more flexible and easier to use than before.Finally, XGBoost Regressor (XGBR) was chosen above Random Forest Regression (RFR) with CatBoost Regression (CR).To put these algorithms to the test in the real world, we will employ a massive dataset that includes data from places like New Delhi, Bangalore, Kolkata, and Hyderabad in particular.We will clean, reduce, and otherwise prepare the dataset to meet our specific requirements before splitting it into a training set and a testing set. In order to get these algorithms into action, we are working hard to make them user-friendly. To find out how effective each method is, we will next compare their performance using various metrics. In doing so, we hope to identify the method for AQI prediction that works best. Its ultimate goal is to unearth intriguing insights about prediction techniques. We will compare the datasets' accuracy levels after using the SMOTE technique to level them. That way, we can see how the imbalanced datasets impact our study. The accuracy of three regression models—Random Forest, XGBoost Regression, and CatBoost—will be assessed using two metrics: R-Square and Root Mean Square Error (RMSE). The models will be tested on both balanced and unbalanced datasets.

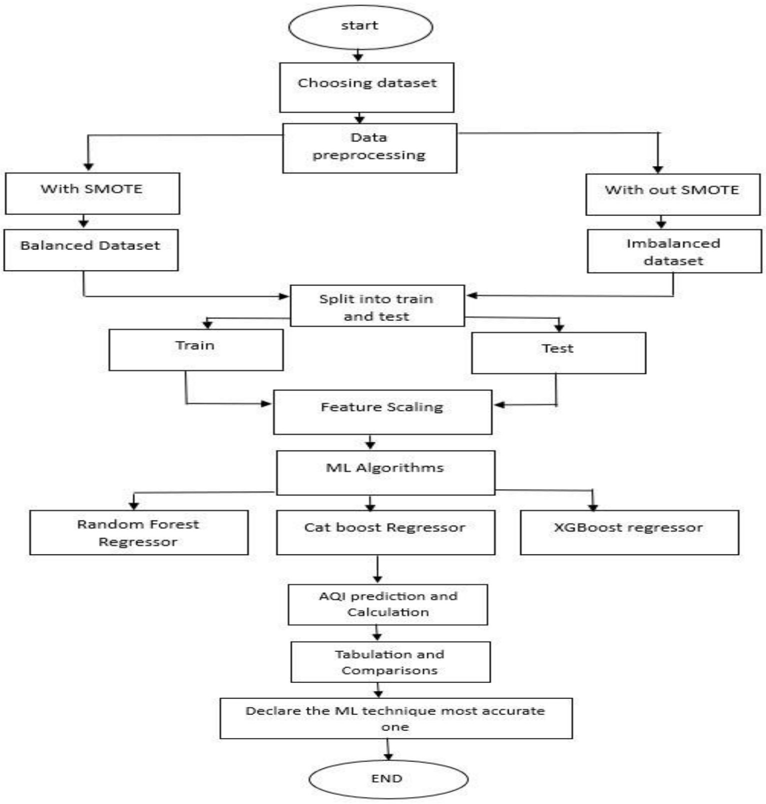

Figure 2: AQI Evolution Process

Finally, the results will be presented comprehensively through clear figures, graphs, and charts, facilitating easy interpretation and paving the way for future research endeavours. We adhered to the methodology outlined in the reference paper, meticulously following each step. The **FIGURE 2** illustrates the sequential process involved in our project development:

AQI Prediction & Calculation of evaluation metric for each ML technique

This method is made possible with the help of machine learning algorithms, which allow for the accurate estimation of AQI for each city. We then create graphs showing the accuracy levels across all eleven cities and tabulate the results. For this project, we will be using R-SQUARE, MSE, and MAE from CatBoost, XGBoost, random forest, and regression. R-Square, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Accuracy are the measures that utilized. **FIGURE 3 and 4** display the actual and predicted values of the AQI using various algorithms.

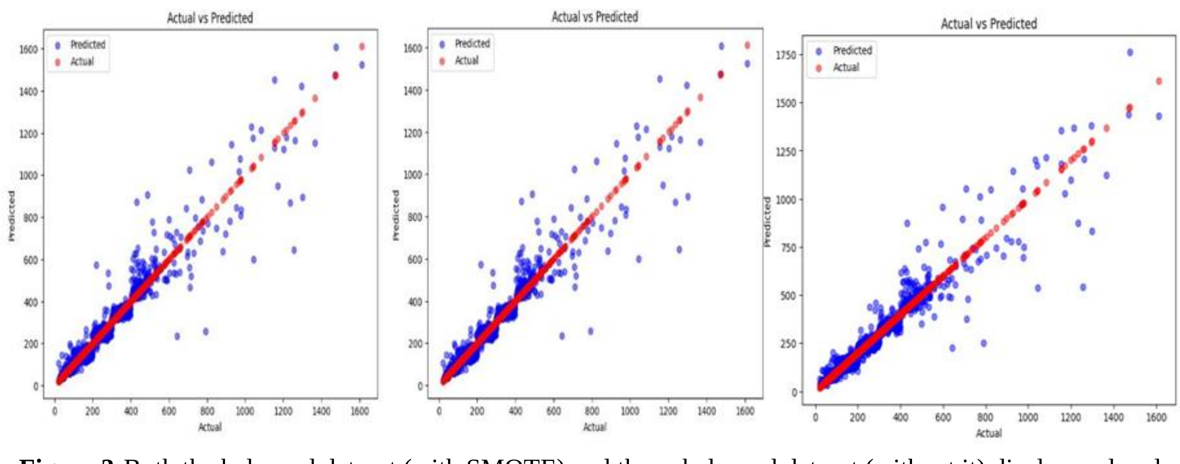

Figure 3: Both the balanced dataset (with SMOTE) and the unbalanced dataset (without it) display real and forecasted AQI values in scatter plots. A Regressor Algorithm Utilizing XGBoost.

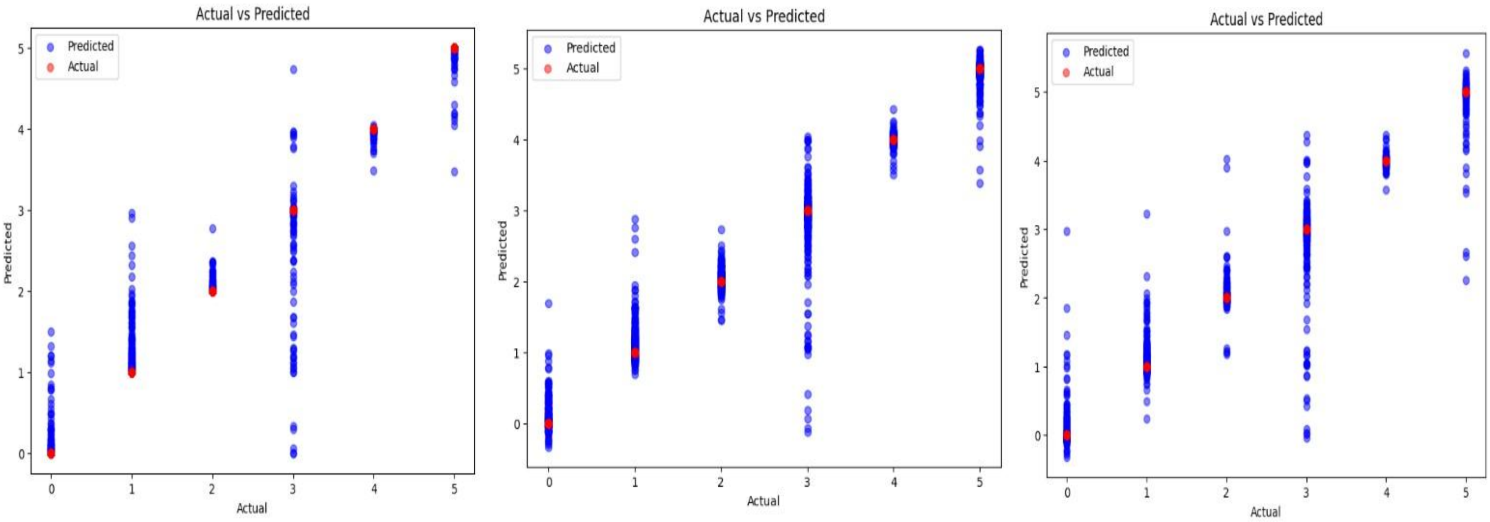

Figure 4: Both the balanced dataset (with SMOTE) and the unbalanced dataset (without it) display real and forecasted AQI values in scatter plots. With the Random Forest Regressor Algorithm.

RESULT AND DISCUSSION

Data related with the Indian cities of Hyderabad, Kolkata, Bangalore, and Delhi were carefully removed in preparation for the proposed study. There were two separate uses for the dataset afterwards: first, with its imbalanced state, and second, after applying SMOTE to make it balanced. The results show that using the balanced dataset significantly improved the models' accuracy. Three algorithms—XGBoost Regression, Random Forest Regression, and Cat Boost Regression—were used for predictive modelling. We offered a visual depiction of the correlation between the experimental data and the anticipated results. For each method, we also determined critical parameters like RSQUARE, MSE, and MAE. The results are presented in detailed tables, graphs, and scatter plots to facilitate your comparison of the balanced and imbalanced datasets. All three methods perform significantly better when given a balanced dataset, as the figures above clearly demonstrate.

CONCLUSION AND FUTURE WORK

Air pollution is a major problem, and scientists throughout the world are scrambling to find a solution. In this paper, we examine the potential use of ML approaches for AQI (Air Quality Index) value prediction. Using air quality index (AQI) data from a number of pollution-prone Indian cities, this paper compares the prediction capabilities of three popular data mining models: XGBoost Regression, Random Forest Regression, and CatBoost Regression. To fix the dataset's class imbalance, we used SMOTE, which stands for Synthetic Minority Oversampling Technique. The results are thus more trustworthy and uniform as a result. As part of our fresh approach, we took a balanced look at the datasets and compared the outcomes from the two sets of data meticulously. When using evaluation metrics like MAE, MSE, and R-SQUARE, it was found that balanced datasets excelled. In the heavily populated urban areas of India, our AQI forecasting models perform far better when using the Synthetic Minority Oversampling Technique (SMOTE), according to our thorough investigation. Random Forest Regression (RFR), CatBoost Regression (93%) and XGBoost Regression (94%) models were able to attain impressive R2 values without SMOTE. But the model's accuracy shot up after SMOTE fixed the dataset's class imbalances. An R2 value of 97% was achieved by the RFR model, 96% by CatBoost, and 96% by XGBoost when SMOTE was employed. The significant improvement in projected accuracy across all three procedures demonstrates the importance of resolving dataset class imbalances. Our research shows that SMOTE can help level datasets, which in turn improves the accuracy of AQI prediction models. The importance of dataset balancing strategies for improving machine learning models' environmental forecasting capabilities is highlighted by the study's findings.

REFERENCES

1. Janarthanan, Ramadoss & Partheeban, Pachaivannan& Somasundaram, S. & Elamparithi, P.. (2021). A Deep Learning Approach for Prediction of Air Quality Index in a Metropolitan City. Sustainable Cities and Society. 67. 102720. 10.1016/j.scs.2021.102720.
2. Londhe, M. 2021. Data Mining and Machine Learning Approach for Air Quality Index Prediction. International Journal of Engineering and Applied Physics. 1, 2 (May 2021), 136–153.
3. Liu, H.; Li, Q.; Yu, D.; Gu, Y. Air Quality Index and Air Pollutant Concentration Prediction Based on

Machine Learning Algorithms. Appl. Sci. 2019, 9, 4069. https://doi.org/10.3390/app9194069

4. Mani, Geetha &Viswanadhapalli, Joshi Kumar & Stonie, Albert. (2021). Prediction and Forecasting of Air Quality Index in Chennai using Regression and ARIMA time series models. Journal of Engineering Research. 9. 10.36909/jer.10253.
5. Baran, Burhan. (2021). Air Quality Index Prediction In Besiktas District By Artificial Neural Networks And K Nearest Neighbors. MühendislikBilimleriveTasarımDergisi. 9. 52-63. 10.21923/jesd.671836.
6. Wang, Jingyang& Li, Xiaolei& Jin, Lukai& Li, Jiazheng& Sun, Qiuhong& Wang, Haiyao. (2022). An air quality index prediction model based on CNN-ILSTM. Scientific Reports. 12. 10.1038/s41598-022-12355-6.
7. Yongliang Feng and Shadi Aljawarneh. 2022. Air Quality Prediction Model Using Deep Learning in Internet of Things Environmental Monitoring System. Mob. Inf. Syst. 2022 (2022). https://doi.org/10.1155/2022/7221157.
8. Maltare, Nilesh & Vahora, Safvan. (2023). Air Quality Index prediction using machine learning for Ahmedabad city. Digital Chemical Engineering. 7. 100093. 10.1016/j.dche.2023.100093.
9. Hansun, Seng &Wicaksana, Arya &Kristanda, M.B.. (2021). Prediction of jakarta city air quality index: Modified double exponential smoothing approaches. International Journal of Innovative Computing, Information and Control. 17. 1363-1371. 10.24507/ijicic.17.04.1363.
10. Jamal A, Nabizadeh Nodehi R. Predicting Air Quality Index Based On Meteorological Data: A Comparison Of Regression Analysis, Artificial Neural Networks And Decision Tree. Japh. 2017;2(1).
11. Ganesh, S. &Pachiyappan, Arulmozhivarman & Tatavarti, Rao. (2017). Forecasting Air Quality Index Using an Ensemble of Artificial Neural Networks and Regression Models. Journal of Intelligent Systems. 28. 10.1515/jisys-2017-0277.
12. Wu, Qunli& Lin, Huaxing. (2019). Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. Sustainable Cities and Society. 50. 101657. 10.1016/j.scs.2019.101657.
13. He, H., & Luo, F. (2020). Study of LSTM Air Quality Index Prediction Based on Forecasting Timeliness. IOP Conference Series: Earth and Environmental Science, 446.
14. Chen, Jeng-Chung & Wu, Yenchun. (2020). Discrete-Time Markov Chain for Prediction of Air Quality Index. Journal of Ambient Intelligence and Humanized Computing. 10.1007/s12652-020-02036-5.