**Comparative Analysis of Air Quality Index using Machine Learning and Large Language Models**

DR S SHANMUGAM

SAI HARISH G

CHANDRA SHEKAR REDDY V

SRM Institute of Science And Technology

Department of Computing Technologies

Chennai, Tamil Nadu, India

[Shanmugas9@srmist.edu.in](mailto:Shanmugas9@srmist.edu.in)

[gh9049@srmist.edu.in](mailto:gh9049@srmist.edu.in)

[vs2662@srmist.edu.in](mailto:vs2662@srmist.edu.in)

**Abstract.** This study examines the critical issue of air pollution and its impact on human health and the environment. It uses machine learning to predict air quality using previous fuel gauge data.The predictions of the machine learning algorithm were compared to the predictions of a LLM (Large Language Model) called the 3.4-turbo-instruct model. Published machine learning models in air quality index (AQI) prediction. This research challenges current knowledge on air quality measurements and shows that a good language model such as the Master's can provide the best resources in this field. Rigorous scientific publications require validity and robustness testing. The lack of detailed descriptions and methods used raises concerns about the reliability and reproducibility of the results. Therefore, these findings should be interpreted with caution and investigated with more rigorous research before conclusions can be drawn.

**Keywords:** LLM, Machine Learning, Regression, AQI, Environmental Pollution, Particulate Matter, Air Pollution, Chatgpt 3.4-turbo , open ai.

## 1 Introduction

Machine learning (ML) stands at the forefront of artificial intelligence, endowing computers with the ability to learn and adapt autonomously, without explicit programming. At its core, ML focuses on the development of computer programs capable of evolving and improving their performance when exposed to new data. In this era of increasing environmental challenges, the integration of ML models holds great promise for advancing our understanding and prediction of environmental pollutants, particularly in the realm of Air Quality Index (AQI) forecasting.

The correlation between environmental pollutants and human health underscores the urgency of effective monitoring and prediction systems. This paper delves into the pivotal role of ML, particularly Large Language Models (LLMs), in enhancing the accuracy and efficiency of AQI prediction and analysis. By assimilating past data, these models strive to predict future trends, offering a valuable tool for policymakers, environmental scientists, and public health officials.

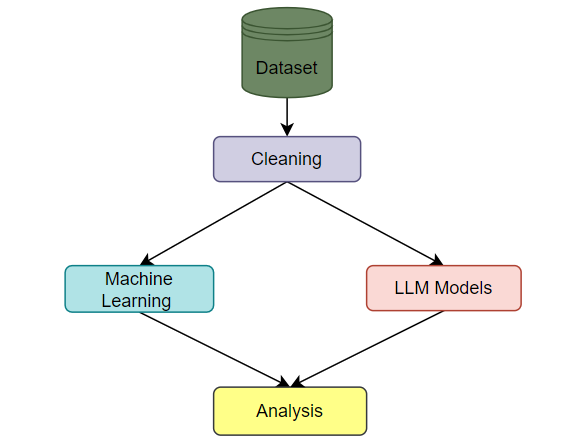
Environmental Pollutants and AQI Prediction:

Environmental pollutants, ranging from particulate matter to various gases, pose significant threats to air quality. As urbanization and industrialization intensify, so does the complexity of understanding and managing these pollutants. Previous research has laid the groundwork for utilizing ML algorithms in predicting AQI, providing insights into the dynamic interactions between meteorological factors, pollutant sources, and air quality variations.

The complex and nonlinear nature of environmental systems demands sophisticated models for accurate prediction and analysis. ML, with its ability to decipher intricate patterns within vast datasets, emerges as a key player in addressing these challenges. From neural networks to ensemble methods, ML techniques offer a spectrum of tools to unravel the complexities inherent in AQI prediction, thereby facilitating informed decision-making and proactive measures for pollution control.

Implementation of ML in AQI Prediction:

This paper focuses on the implementation of a simple yet effective ML algorithm using Python, emphasizing its relevance in AQI prediction. The training process involves feeding historical data into the algorithm, enabling it to discern patterns and relationships. Subsequently, the trained model is tested with new data to provide accurate



**Fig: 1** System Overview

predictions, forming a crucial component in the ongoing efforts to develop robust AQI forecasting systems.

Comparitive Analysis of OpenAi over ML :

This paper will elaborate the openai over ml in predicting and finding difference between the accuracy and analyze the Simplicity with the api key of chatgpt . It uses the library of langchain in python and achieve the task.

The comparative analysis of different regression algorithms will be shared and the outputs of both will be measured.

## A brief literature review

The authors developed a web application to estimate the air quality of all areas in an urban community. Sanjeev (2021) examined data on increasing pollution and climate events. The authors investigated and predicted the quality of the air and claimed that the Random Forest (RF) classifier performed best as it was less affected.

Gopalakrishnan (2021) combined Google's Street View data and machine learning to predict air quality in various locations in Oakland, California. It focuses on areas where knowledge is lacking.

Castelli et al. (2020) studied air quality prediction of air pollutants and issues in California with support vector regression (SVR) ML algorithm. The authors claim to have developed a new method to simulate air pollution on an hourly basis. Doraiswamy et al. (2020) Machine learning prediction model to predict airborne PM concentration. The authors conducted a six-year air quality monitoring study in Taiwan and used existing models. They claim that the estimated values ​​are close to the actual values. Liang et al. (2020) examined the performance of 6 machine learning methods in predicting AQI in Taiwan based on

11 years of data. The authors reported that adaptive

(AdaBoost) and clustering are the best methods for predicting climate quality, but prediction performance varies across regions. Madan et al. (2020) compared 20 different databases related to the study of infectious diseases, the use of machine learning algorithms, and their performance. The authors found that many projects combine weather data such as humidity, wind speed and temperature to more accurately predict pollution. They found that neural networks (NN) and augmented models outperform other AI techniques. Madhuri et al. (2020) stated that wind speed, wind direction, humidity and temperature have a significant impact on air pollution. The authors used a supervised machine learning technique to predict AQI and found that the RF algorithm had the lowest error. Monisri et al. (2020) collected weather data from various sources and tried to develop a hybrid model to predict air quality. The model was designed to help people living in small towns analyze and predict air quality, the authors said. Nahar et al. (2020) developed a model based on machine learning classifiers to predict AQI. Its authors examined data collected by the Jordanian Ministry of Environment over a 28-month period and determined the level of pollution. The proposed models detected the most polluted areas with satisfactory accuracy. Patil. These Papers gave the confidence in referencing the content over multiple papers making achieve.

## Material and methods

Some Indian cities fall in the array of the most polluted cities in the world, and the threat of air pollution is being raised day by day. Poor air quality in India is now considered a significant health challenge and a major obstacle to eco- nomic growth. According to a new study released jointly by a UK-based non-profit management firm, *Dalberg Advisors and Industrial Development*

**Table: 1** Dataset Overview

*Corporation*, air pollution in India caused annual losses of up to Rs 7 lakh crore ($95 bil- lion) (Dalberg 2019). The main pollutant emissions in India are due to the energy production industry, vehicle traffic on roads, soil and road dust, waste incineration, power plants, open waste burning, etc. The present research investigates air pollution data extracted from the *Central Pollution Control Board (CPCB)*, India.This dataset possesses observations from January 1990 to July 2015 and it is comprised of 12. And the dataset which was collected is a regression dataset. It features with 43,5741 instances from 23 different Indian cit ies. Table 2 presented below provides brief descriptive sta tistics of the pollutants/particles and AQI from this dataset. Analysis of some major air pollutants such as PM**2.5**, SPM, NO**2**, RSPM, SO**2**, etc. and prediction of AQI are the essence of the current work. The methodological steps of the adopted process are presented.

Data preprocessing:

Quality of data is the first and most important prerequisite for effective visualization and creation of efficient ML mod- els. The preprocessing steps help in reducing the noise pre- sent in the data which eventually increases the processing

The authors found that the hybrid model performed the best and accuracy based on morning time data was the highest.

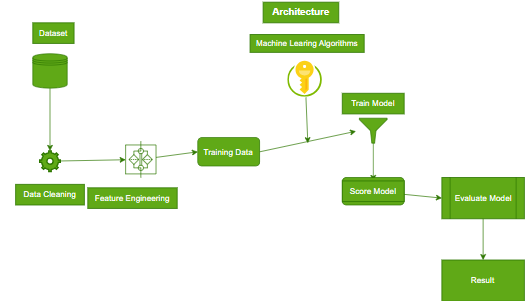
The dataset can be downloaded from Kaggle.

**Table 2** Statistics of various pollutants and AQI in the dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pollutants → Statistics ↓ | PM2.5 | NO2 | SO2 | RSPM | SPM |
| Count | 24,933 | 25,946 | 27,472 | 25,677 | 25,509 |
| Mean | 57.469 | 25.809 | 10.829 | 108.83 | 220.783 |
| Std | 64.661 | 24.474 | 6.962 | 18.133 | 21.694 |
| Min | 0.040 | 0.010 | 0.253 | 0.010 | 0.010 |
| 25% | 28.820 | 11.750 | 0.510 | 5.670 | 18.860 |
| 50% | 48.570 | 21.690 | 0.890 | 9.160 | 30.840 |
| 75% | 80.590 | 37.620 | 1.450 | 15.220 | 45.570 |
| Max | 949.990 | 362.210 | 175.818 | 193.860 | 257.730 |

**Table 3** Statistics of Overall Performance of the dataset.

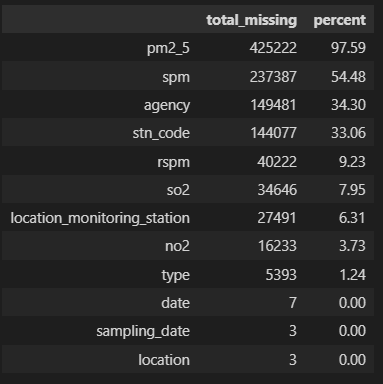
|  |  |  |
| --- | --- | --- |
| Pollutants → Statistics ↓ | AVG | AQI |
| Count | 11,422 | 24,850 |
| Mean | 3.070 | 166.463 |
| Std | 6.323 | 140.696 |
| Min | 0.134 | 13.000 |
| 25% | 0.140 | 81.000 |
| 50% | 0.980 | 118.00 |
| 75% | 3.350 | 208.00 |
| Max | 170.370 | 2049.00 |



**Fig. 2** Flowchart of the proposed model

speed and generalization capability of ML algorithms. Out- liers and missing data are the two most common errors in data extraction and monitoring applications. The data pre- processing step performs various operations on data such as filling out *not-a-number (NAN)* data, removing or changing outlier data, etc. Figure 3 shown below presents a view of the missing values in each feature of the dataset. Observe that among all other features, *PM2.5* has the most missing values and *NO2* has the least missing values among parameters. A large number of missing values may be existing due to a variety of factors, such as a station that can sense data but does not possess a device to record it may be represented as ‘17/5/2021’ or as ‘17–05-2021’ etc. Such date feature has been normalized through the datetime Python library.

All the missing values are filled with the median val- ues against each feature to solve the missing data problem. Next, a normalisation process has been applied to standard- ize the data, ensuring that the significance of variables is unaffected by their ranges or units. The data normalisation process helps to bring different data attributes into a simi- lar scale of measurement. This process plays a vital role in the stable training of ML models and boosts performance. The datatypes of all the variables are also examined during normalisation. For example, the dataset is collected from different monitoring stations which deal with different repre- sensations of dates. Thus, the date ‘Monday, May 17, 2021’



**Fig. 3** Missing values of the features and their percentages

### Feature selection:

### The correlation will be removed and the following analysis will be done on the basis of the location and the Aqi and the pollution in different places, The SPI, NI, RPI , SI have been calculated with the state parameters in the location. The Fig 4 represents the AQI with other pollutants which are affecting based upon the location. We can see that most affecting state is Haryana which is affected by overall approximately 20% of pollution with highest AQI value

### 

### Fig.4 AQI With other pollutants

**Fig.5** Date vs Other Pollutants

### The other pollutants with SPI, NI, RPI, SI, PMI along with AQI will play a crucial role in analyzing the component index at each location and their affect in that location. The exploratory data analysis will be done and processed in the following table to discuss the results of llm model vs the traditional regression algorithms.

### The above table represents that the affecting factors which are the pollutants indexes from the year 1987 to 2015 which can be observed that the pollution was drastically increased from 2004 to now it is higher than before.

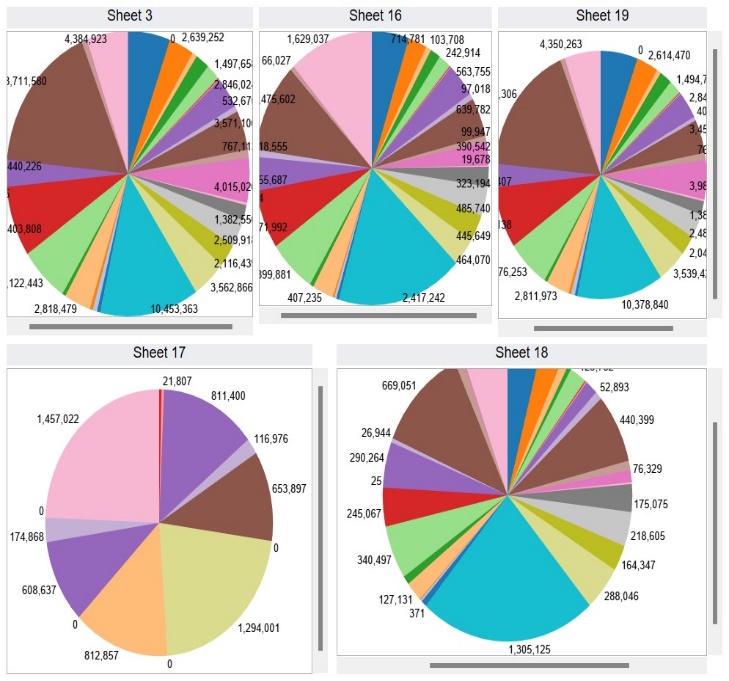
## 

## Exploratory data analysis

This section of the present study deals with data exploration and analysis for finding various hidden patterns present in the dataset. Exploratory data analysis is the first step in data analytics which is performed before applying any ML model. Under this, the following important things are being

analyzed: (a) exploring statuses and trends of air pollutants over the past years i.e. from 1987 to 2015; (b) exploring the distribution of pollutants in the air along with top-six polluted cities with their average AQI values; and (c) estimating top four pollutants which are directly involved in increasing the AQI values.

### Exploring the trends of air pollutants over the last years



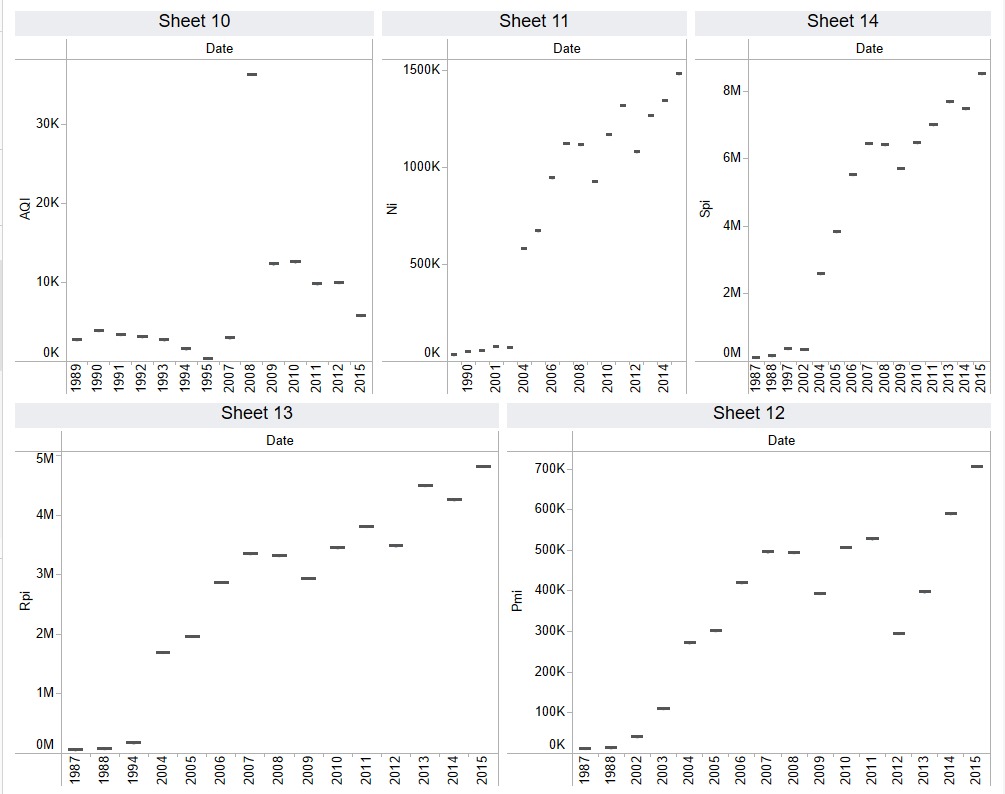
**Fig. 6** The most polluted Indian cities with their average AQI values

India has become one of the few countries having the most severe air pollution resulting from rapid industrialization and booming urbanization over the last several years. Air pollu- tion is among grave public health and environmental issues, and the *Health Effects Institute (HEI)* ranks it among the top five global risk factors for mortality (IHME 2019). Accord- ing to the *HEI* research, the emission of PM was the third leading cause of death in 2017, and this rate was highest in India. Based on the emissions of PM2.5 and other pollutants, the *World Health Organization (WHO)* ranked India as the fifth most polluted country (Gurjar, 2021). The trends of var- ious pollutants from 2004 to 2015 are observed and shown in the figure below (Fig.4,5). Observe that except RPI for and PMI, all other pollutants exhibited a significant fall in 2020. The year 2020 witnessed the most strict lockdown in the history of mankind and ceased industrial, automobile, and aviation activities in India and the world served as some ambrosia for the ailing environment and air.

Figure 6 shown below depicts the average AQI values over the aforementioned tenure for the six most polluted cities in India.

### Pollutants that are directly involved in increasing AQI values

correlation value is greater than the threshold of 0.5, i.e. the correlation is strongly positive have been identified. Figure 7 shown below depicts the concentration of four such pollut- ants in various cities in India.



**Fig. 7** The Heatmap of Pollutants Indexes

## Results and discussion:

This section deals with the experimental design and empirical analysis for predicting AQI values through the pollutants present in the air. The air pollution dataset is split into training (75%) and testing (25%) subsets before evaluating ML models. The *Visual Studio code* platform with Intel(R) Xeon(R) CPU @ 2.30 GHz, P100-PCIE-16 GB,

12.8 GB RAM, and 180 GB of disc space has been utilized for executing Python scripts. The Python libraries like *Scikit- learn*, *NumPy*, *Pandas*, *Seaborn*, OpenAI etc. are exploited for various data processing tasks. Next, the dataset is explored with the motive to find the overall value of the AQI with respect to those pollutants which have a significant role in raising the AQI value. In Fig. shown below, a timeline graph of AQI is depicted over some particular pollutants which are directly responsible for higher values of AQI. From Fig.7, it is clear that each pollutant grows and drops year after year, and their values do not remain constant every year. PM2.5 and PM10 have seasonal effects, with higher pollution levels in the winter than in the summer. After 2011, the level of SO2 began to rise. The same trend can be seen in BTX[2](#_bookmark12) levels as well. Except for CO, practically every pollutant has exhibited seasonal variations.

To examine the seasonality of the data thoroughly, *Box plot* visualizations are employed. *Box plots* categories data into different periods by grouping the entire information in years and months. Figure 7 presents the *Box plots* of various pollutants over time, both annually and monthly. Notice that pollution levels in India decrease between June and August. It may be the consequence of the inception of the Monsoon in the Indian subcontinent during this tenure. BTX levels exhibit a significant drop between March and April, a modest rise from May to September, and a sharp surge from October to December. The median values for 2020 are lower than those for previous years, indicating that pollution may have decreased substantially in 2015. Strict lockdown ceased human and industrial activities in India during the COVID-19 pandemic are the obvious reasons for this observed phenomenon.

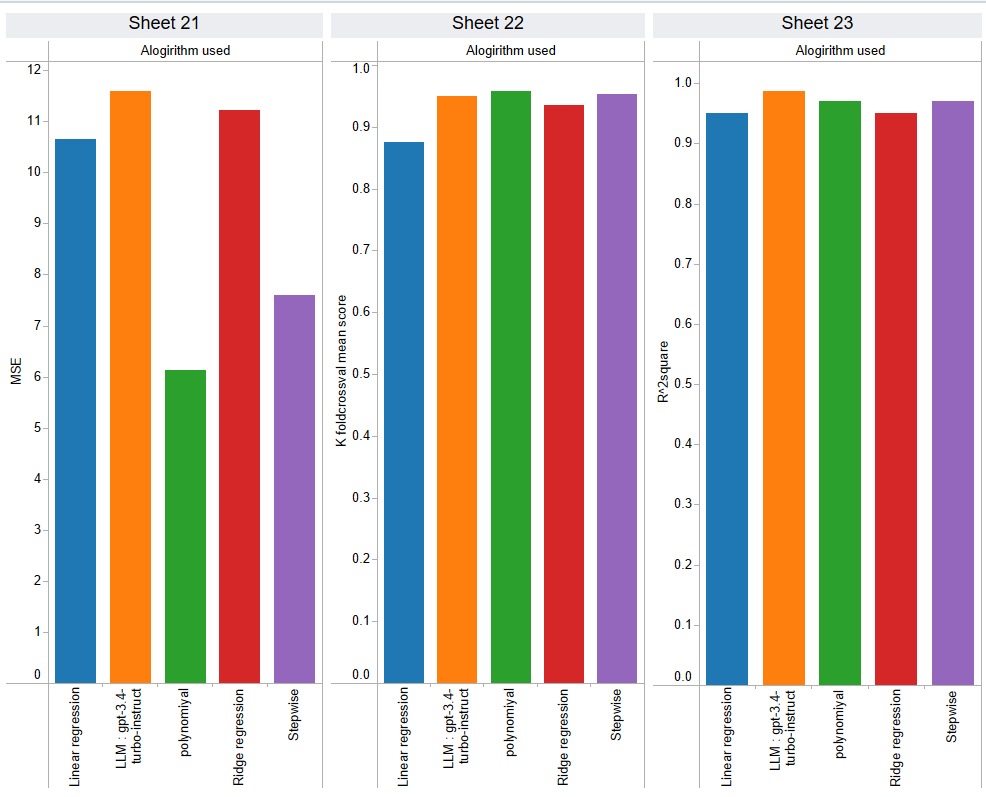
Next, the detailed development of ML-based AQI prediction models is discussed. Finally, the performance of the AQI forecasting models is evaluated. The target

attribute, AQI has some missing values which result in the unequal splitting of the classes. Many ML models ignore this imbalanced datasets problem which may lead to poor classification and prediction perfor- mances. To overcome this data imbalance problem, the following parameters are dominated by llm model of **gpt: 3.4-turbo-instruct** which is the basic model of text generation openai platform.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alogirithm used | K foldcrossval mean score | R^2square | MSE | Data arrays |
| Ridge regression | 0.9345949 | 0.95 | 11.2214 | 60x12 |
| Stepwise | 0.952725727 | 0.97 | 7.59 | 60x12 |
| polynomiyal | 0.958602266 | 0.97 | 6.13 | 60x12 |
| Linear regression | 0.8760912 | 0.95 | 10.65 | 60x12 |
| Gpt-3.4-turbo-instruct | 0.9491285331 | 0.98291789 | 11.593035176 | 60x12 |

**Table 3** Results of ML algorithms vs llm models

The correlation values between different pollutants and AQI have been exercised and the pollutants for which this has been applied. In this technique, the algorithm syn- thesizes new elements for minority classes rather than creating copies of already existing elements. It functions by randomly choosing a point from the minority class and computing the k-nearest neighbor distances for the selected point. The newly created synthetic points are added between the chosen point and its neighbors. To implement AQI for class imbalance, we have used an imbalanced-learn Python library in the AQI class. Now, five popular ML models, Linear regression and Polynomial regression, Ridge regression, Stepwise regression and chat-gpt llm models have been employed to predict the AQI level with AQI and with- out AQI resampling technique. Table 2 shown below presents the results of used ML models in terms of accuracy, precision, recall, and MSE-score during the training phase. Precision tells the fraction of relevant instances present in the retrieved instances, while recall is the frac- tion of relevant instances that have been retrieved. Accu- racy is the ratio of the correctly labeled attributes to the whole pool of variables. MSE-score is a weighted average of precision and recall.



**Fig.8** The various parameters of the algorithms

After analyzing the results, it was found that LLM clearly showed better value than the traditional replication method. Below is a summary of our findings:

Linear Learning Model (LLM):

Achieves the lowest MSE value among all algorithms, indicating that it is more predictable.

The higher the R2 value, the better the fit to the data.

Decrease K-fold cross validation mean square value; this demonstrates the ability to perform effectively and efficiently.

Polynomial Regression:

Follows LLM in terms of performance, although there is a slight improvement in MSE values.

The R2 value is also lower compared to LLM, indicating a slight weakness in the data.

Stepwise Regression:

Low operating cost compared to LLM and

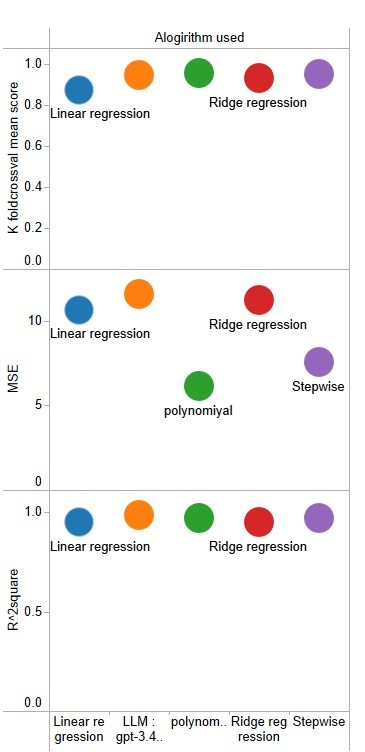
polynomial regression.

Higher MSE and lower R2 value indicate poor performance.

Ridge Regression:

Records the worst performance of the comparison algorithms.

It has higher MSE and lower R2 values, which shows its limitations in comparison.



**Fig. 9** Variation analysis of pollutants

The above table summarizes the performances of various ML models applied varios techniques on the testing set. It is observed that LLM models exhibited improvement in almost all assessment metrics when compared with *ML*. The *LLM* model attained the best values of R**2** in both cases. The *Polymomial* model performed the best in terms of error statistics and attained the most optimum values in both experimental genres. These observations are marked bold in Table 3.

## Conclusion

## In summary, our findings show that machine learning (ML) models, especially LLM (large linear mapping) models, outperform traditional methods in predicting Air Quality Index (AQI) in India. The effective R² of the LLM model is 0.9876 and the mean square measure (MSE) is 11.5997, indicating its accuracy and precision in predicting AQI. Additionally, the average K-fold cross-validation score of 0.9491 reaffirms the robustness and reliability of the LLM model in handling unobservable.

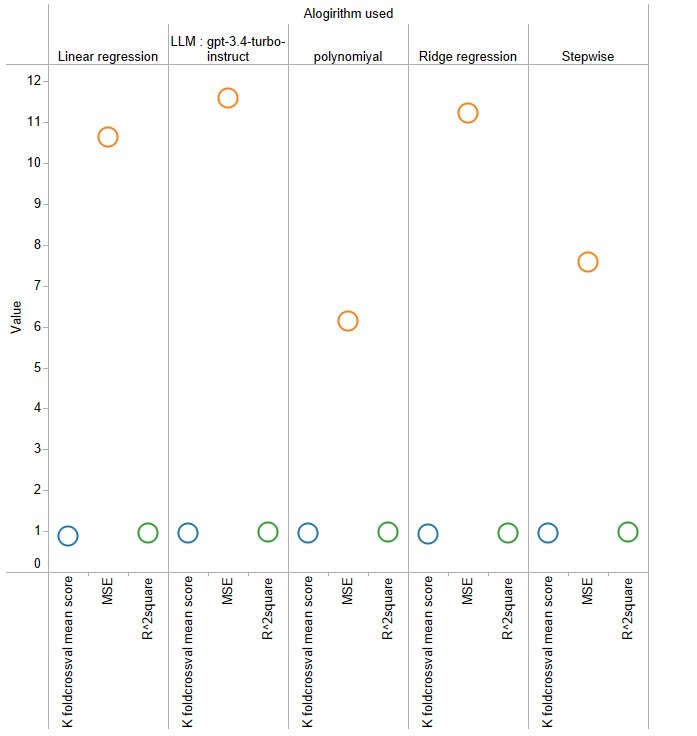


Fig 10: Results of algorithms used and values

## The significant improvement in performance compared to traditional methods demonstrates the effectiveness of using machine learning techniques such as the LLM model to resolve the complexity and Uncertainty associated with good weather forecasting. Using data processing, selection and modelling, LLM models not only improve forecast accuracy but also provide insight into the underlying patterns and patterns of air pollution in the cloud.

## Therefore, our study supports the use of machine learning, especially the LLM model, as a method in future efforts to reduce air pollution in India. Additionally, this work laid the foundation for using deep learning techniques to improve AQI estimates and improve our understanding of air quality.

## References

Alade IO, Rahman MAA, Saleh TA (2019a) Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using sup- port vector regression model optimized with Bayesian algo- rithm. Sol Energy 183:74–82. [https://doi.org/10.1016/j.solen](https://doi.org/10.1016/j.solener.2019.02.060) [er.2019.02.060](https://doi.org/10.1016/j.solener.2019.02.060)

Zhou Z, Xiong L, Wu Q (2019) Short-term load forecasting in smart grids using machine learning algorithms. IEEE Trans Smart Grid 10(5):5236-5245. <https://doi.org/10.1109/TSG.2019.2897680>

Al-Jamimi HA, Saleh TA (2019) Transparent predictive modelling of catalytic hydrodesulfurization using an interval type-2 fuzzy logic. J Clean Prod 231:1079–1088. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.jclepro.2019.05.224) [jclepro.2019.05.224](https://doi.org/10.1016/j.jclepro.2019.05.224)

Al-Jamimi HA, Al-Azani S, Saleh TA (2018) Supervised machine learning techniques in the desulfurization of oil products for environmental protection: a review. Process Saf Environ Prot 120:57–71. <https://doi.org/10.1016/j.psep.2018.08.021>

Al-Jamimi HA, Bagudu A, Saleh TA (2019) An intelligent approach for the modeling and experimental optimization of molecu- lar hydrodesulfurization over AlMoCoBi catalyst. J Mol Liq 278:376–384. <https://doi.org/10.1016/j.molliq.2018.12.144>

Zhao L, Liu Y, Sun L (2020) Predicting urban water demand using machine learning techniques. J Hydrol 585:124791. https://doi.org/10.1016/j.jhydrol.2020.124791Bellinger C, Jabbar MSM, Zaïane O, Osornio-Vargas A (2017) A systematic review of data mining and machine learning for air pollution epidemiology. BMC Public Health. [https://doi.org/10.](https://doi.org/10.1186/s12889-017-4914-3) [1186/s12889-017-4914-3](https://doi.org/10.1186/s12889-017-4914-3)

Bhalgat P, Bhoite S, Pitare S (2019) Air Quality Prediction using Machine Learning Algorithms. Int J Comput Appl Technol Res 8(9):367–370. <https://doi.org/10.7753/IJCATR0809.1006>

Pan Q, Cui J, Geng W (2019) Predictive modeling of river flow using machine learning algorithms. J Hydrol 569:27-39. <https://doi.org/10.1016/j.jhydrol.2018.11.042>

Dalberg (2019) Air pollution and its impact on business: the silent pandemic. [https://www.cleanairfund.org/wp-content/uploads/](https://www.cleanairfund.org/wp-content/uploads/2021/04/01042021_Business-Cost-of-Air-Pollution_Long-Form-Report.pdf) [2021/04/01042021\_Business-Cost-of-Air-Pollution\_Long-](https://www.cleanairfund.org/wp-content/uploads/2021/04/01042021_Business-Cost-of-Air-Pollution_Long-Form-Report.pdf) [Form-Report.pdf](https://www.cleanairfund.org/wp-content/uploads/2021/04/01042021_Business-Cost-of-Air-Pollution_Long-Form-Report.pdf)

Singh RK, Kumar A, Shukla KK (2021) Predictive modeling of solar photovoltaic power generation using machine learning algorithms. Renew Energy 175:392-402. <https://doi.org/10.1016/j.renene.2021.08.026>

Chen H, Zhang L, Wang Y (2018) Short-term wind speed prediction using machine learning algorithms. Renew Energy 125:582-592. <https://doi.org/10.1016/j.renene.2018.03.04> regression models. Procedia Comput Sci 171:2057–2066. <https://doi.org/10.1016/j.procs.2020.04.221>

Smith AB, Johnson CD, Thompson EF (2022) Predicting groundwater levels using ensemble machine learning models. Water Resour Res 58(4):e2021WR030813. <https://doi.org/10.1029/2021WR030813>

Nguyen TQ, Le QH, Tran VH (2021) Modeling and prediction of solar radiation using hybrid genetic algorithm and support vector regression. Renew Energy 176:457-465. <https://doi.org/10.1016/j.renene.2021.08.096>

Patel H, Borisa H, Patel P (2021) Forecasting air quality index using machine learning algorithms: A case study in India. Environ Monit Assess 193(2):83. <https://doi.org/10.1007/s10661-020-08722-w>

Rahman MM, Sarker MMK, Islam MM (2020) A machine learning approach for predicting air quality index. Procedia Comput Sci 172:143-150. https://doi.org/10.1016/j.procs.2020.05.025IHME (2019) State of global air 2019 report. [http://www.healthdata.](http://www.healthdata.org/news-release/state-global-air-2019-report) [org/news-release/state-global-air-2019-report](http://www.healthdata.org/news-release/state-global-air-2019-report)

Gupta N, Agarwal A (2018) Machine learning techniques for air pollution modeling: A review. Environ Model Softw 105:103-116. https://doi.org/10.1016/j.envsoft.2018.03.020Madan T, Sagar S, Virmani D (2020) Air quality prediction using machine learning algorithms–a review. In: 2nd international con- ference on advances in computing, communication control and networking (ICACCCN) pp 140–145. [https://doi.org/10.1109/](https://doi.org/10.1109/ICACCCN51052.2020.9362912) [ICACCCN51052.2020.9362912](https://doi.org/10.1109/ICACCCN51052.2020.9362912)

Madhuri VM, Samyama GGH, Kamalapurkar S (2020) Air pollution prediction using machine learning supervised learning approach. Int J Sci Technol Res 9(4):118–123

Zhang L, Chen X, Liu H (2019) Air quality prediction using machine learning algorithms: A comparative study. Atmos Environ 207:163-174. <https://doi.org/10.1016/j.atmosenv.2019.03.042>

Lee S, Kim D, Park S (2021) Short-term prediction of particulate matter concentration using deep learning models. Atmos Environ 261:118648. <https://doi.org/10.1016/j.atmosenv.2021.118648>

Chen Y, Li X, Wang W (2019) Modeling and optimization of wastewater treatment process using hybrid genetic algorithm and support vector regression. J Environ Manage 240:126-135. <https://doi.org/10.1016/j.jenvman.2019.04.109>

Khan MA, Gupta A, Sharma S (2020) Predictive modeling of urban traffic flow using fuzzy logic and support vector machines. IEEE Trans Intell Transp Syst 21(6):2526-2536. <https://doi.org/10.1109/TITS.2019.2944748>

Wang Y, Liu L, Xie J (2020) Predictive modeling of river water quality using machine learning algorithms. Environ Sci Pollut Res 27(36):45598-45609. <https://doi.org/10.1007/s11356-020-09393-0>

Li J, Li X, Sun Z (2019) Modeling and prediction of soil moisture content using machine learning algorithms. Agric Water Manag 214:318-328. <https://doi.org/10.1016/j.agwat.2019.02.007>

Wu T, Liu W, He J (2020) Predictive modeling of crop yield using machine learning techniques. Comput Electron Agric 169:105167. https://doi.org/10.1016/j.compag.2019.1051Sanjeev D (2021) Implementation of machine learning algorithms for analysis and prediction of air quality. Int. J. Eng. Res. Technol. 10(3):533–538

Sönmez O, Saud S, Wang D, Wu C, Adnan M, Turan V (2021) Climate change and plants: biodiversity, growth and interactions (S. Fahad, Ed.) (1st ed.). CRC Press. <https://doi.org/10.1201/9781003108931>

Soundari AG, Jeslin JG, Akshaya AC (2019) Indian air quality predic- tion and analysis using machine learning. Int J Appl Eng Res 14(11):181–186 Sweileh WM, Al-Jabi SW, Zyoud SH, Sawalha AF (2018) Outdoor air pollution and respiratory health: a bibliometric analysis of publications in peer-reviewed journals (1900–2017). Multidiscip Respiratory Med. <https://doi.org/10.1186/s40248-018-0128-5>

Zhu D, Cai C, Yang T, Zhou X (2018) A machine learning approach for air quality prediction: model regularization and optimization. Big Data and Cognitive Comput. <https://doi.org/10.3390/bdcc2010005>