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

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 pollution of air and it has impact on to the environment and the human health. It uses machine learning which will be used 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 a great 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.

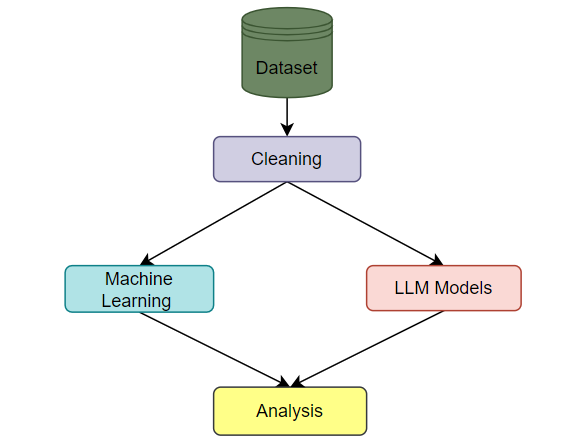
## 1 Introduction

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.

**1.1 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.

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

**1.3 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.

**Fig: 1** System Overview

## 2 A brief literature review

The authors developed web application to measure air quality in all areas of urban communities. Sanjeev (2021) reviews the literature on air pollution and climate. The authors predicted and analyzed air quality finally found that the Random Forest which is (RF) classification performed very well due to its smaller impact. , California. It focuses on areas where knowledge is lacking. (2020) studied the air quality prediction of air pollution and the California problem using support vector regression (SVR) ML algorithm. It’s Authors claim to have developed a new method to simulate hourly weather. Doraswamy et al. (2020) Machine learning prediction models which can be Regression to predict atmospheric PM concentration. The authors conducted a six-year air quality study in Taiwan and used existing models. They claim that the estimate is close to the true value. Liang et al. (2020) examined the performance of 6 ml methods in predicting AQI in Taiwan. The authors report that adaptive

(AdaBoost) and clustering are the best methods for predicting climate quality, but prediction performance varies by region. Madan et al. (2020) used machine learning algorithms to compare 20 different databases on infectious disease detection and performance. The authors found that many projects combine weather related information such as wind speed, humidity and temperature so that the pollution will be predicted more accurately. In neural networks (NN) they found that and continuous models outperform other AI methods. Some author found that wind direction, wind speed, temperature, and humidity have significant effects on climate. The authors used supervised ML to predict air quality index and discovered that RF algorithm had the lowest error. Monistic et al. The authors say the model was designed to help small-town residents analyze and predict air quality. Dinner and other things. (2020) developed an AQI prediction model based on machine learning classification. The authors analyzed data collected by the Jordanian Ministry of Environment over a 28-month period and determined the prevalence of the disease. The proposed model identified the most polluted areas with satisfactory accuracy. Patil's article provides a reliable reference to various documents to understand the main points.

## 3 Material and methods

Some of Indian cities are under the threat of air and environmental pollution which is increasing gradually. Among the most polluted cities in world, and India's air pollution is now seen as a major health problem and major impediment to economical growth. That's according to new research published jointly by Dahlberg Consultants and the British non-profit Industrial Development Corporation.

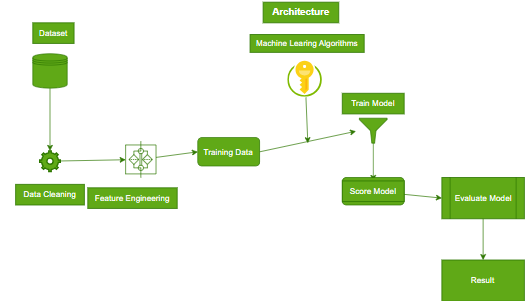
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pollutants → Statistics ↓ | PM2.5 | NO2 | SO2 | RSPM | SPM |
| Count | 24,933 | 25,946 | 27,472 | 25,677 | 25,509 |
| Mean | 57.469 | 25.809 | 10.829 | 108.83 | 220.783 |
| Std | 64.661 | 24.474 | 6.962 | 18.133 | 21.694 |
| Min | 0.040 | 0.010 | 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 1** Statistics of different Pollutants and AQI in Dataset

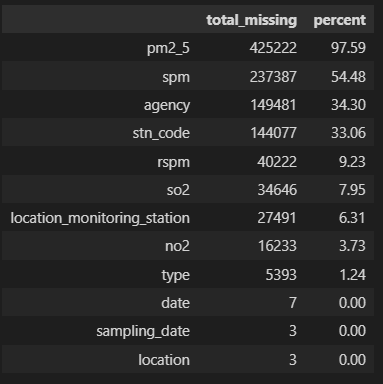
The company estimates that the cost of pollution of air in India is Rs. Loss of up to 7 billion rupees ($95.5 billion) (Dalberg 2019). The major pollution in India comes from the energy industry, traffic, waste incineration, road dust, open waste burning, power plants etc It is due to. (CPCB), India. The database contains observational data from January 1990 to July 2015 and contains 12 files called regression datasets. 43,5741 cases are reported from 23 of Indian cities different. Table 2 provides a brief description of pollutant/minor, AQI in this databases. PM2.5, SPM, NO2, RSPM, SO2, analysis of these pollutants and estimation of AQI are main focus of the current study. The standard procedure of the adopted procedure is presented. According to the company, the cost of air pollution in India is up to 7 billion rupees ($95 billion) every year (Dalberg 2019). The major pollution in India is energy generation, traffic, land and waste incineration, road dust, open waste burning, ,power plants etc. due to reasons. This database contains survey data from January 1990 to July 2015 and contains 12 documents. There are 43,5741 cases from 23 various cities in India. Table 2 provide a brief description of the contaminant/minor and air quality index in this database. Analysis of few pollutants like PM2.5, SPM, SO2, NO2, RSPM and estimation of AQI are focused mainly of current study. The standard procedure of the adopted procedure is presented.

**3.1 Data preprocessing:**

The dataset can be downloaded from Kaggle.



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

Generalization and speed capabilities of machine learning algorithms. Missing data and Outliers are two of the errors which are most common in data collection and operations monitoring. Data processing procedures on data, such as populating non-numeric data (NAN), extracting or modifying other data. Figure 3 below shows the types of missing values ​​for each database. Note that among other characteristics, PM2.5 is missing values ​​and NO2 has the fewest number of values. Sites can recognize the data but don't have the tools to save it, "11/5/2021" or "11-05-2021" etc. There may be many missing values ​​for various reasons such as: library day. To solve the problem of missing data, the mean value of each feature is calculated. The data is then normalized using the normalization process to ensure that changes do not affect quantity or unit values. Data processing techniques help combine disparate data into similar measurements. This process plays an important role in training stable learning models and improving their performance. Information on each variable is also checked during the validity period. For example, a set of data collected from different monitoring stations, in different situations, in different formats and on different dates. Therefore, the date is "Monday, May 11, 2021".

### 3.2 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. 3** Missing values of the features and their percentages

### 

### 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.

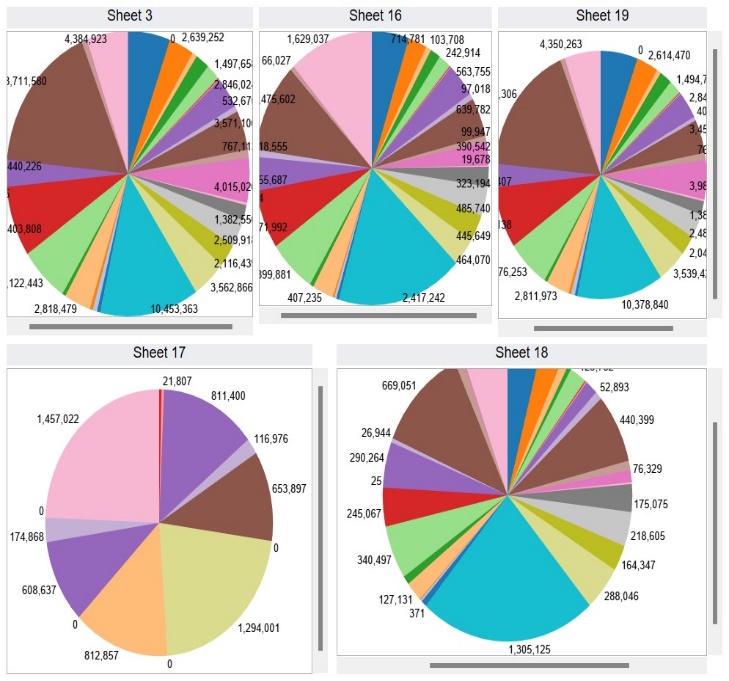
## 

## 4 Exploratory data analysis

Part of the current study involves searching and analyzing data to discover several key patterns present in literature. Data analysis is first step in data analysis and is done before applying machine learning models. The following main items are mentioned: (1) Research on climate change and trends in recent years (i.e, from 1987 to 2015); (b) Distribution of air pollution and evaluation of air distribution in the most polluted cities Average Air Quality Index; (c) Evaluate 4 pollutants that are directly associated with increasing air quality index values. And Exploring of air pollution trends in recent years:

Based upon the location. We can see that the most affected state is Haryana which is affected by overall approximately 20% of pollution with the highest AQI value

**4.1 Exploring air pollutants of different trends over last few years**



**Fig. 6** Most polluted Indian cities with their AQI values

India has become one of the least polluted countries in the recent year due to rapid increases in industrialization and booming the urbanization. Air pollution is the most one and important public health and environmental problems and is one of five leading which cause of death in world based up on the Health Effects Institute (HEI) (IHME 2020). According to HEI research, PMI is the third leading cause of death, the highest in India in 2017. The (WHO) has identified India as the 5th most polluted country in terms of PM2.5 emissions and other pollutants (Gurjar, 2021). Changes in various pollutants from 2004 to 2015 are shown in figure below (Figure 4). All pollutants except RPI and PMI see declines in 2020. Environment and security.

### 4.2 Pollutants directly involved in increasing Air Quality Index values

A correlation value greater than 0.5 means that the correlation is defined as strong positive. Figure 7 below is up on the concentration of 4 such pollutants in various cities of India.

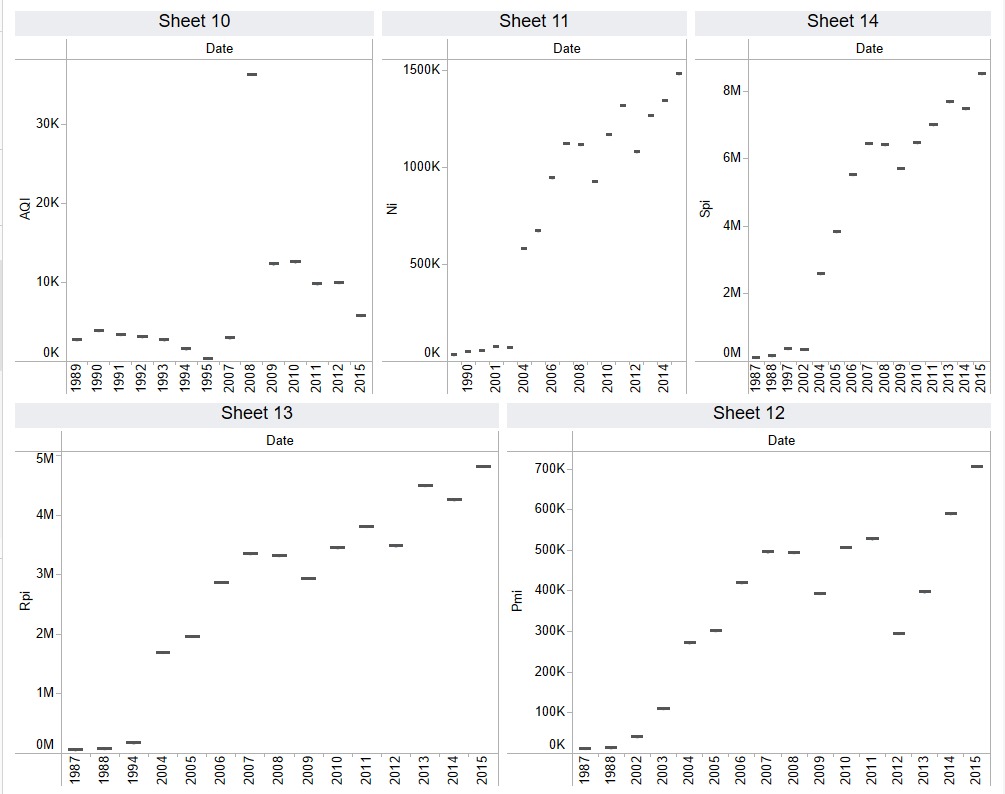


Fig. 7 The Heatmap of Pollutants Indexes

## 

## 5 Final Discussion and Results:

This section discusses about analysis and experimental design of using air pollutants to estimate AQI values. Before evaluating the ML model, meteorological data were divided into training (75%) and testing (25%). Create Python scripts using the Visual Studio Code platform with Xeon (R) Intel (R) CPU @ 2.40 GHz, P200-PCIE-32 GB,

14.8 GB RAM and 150 GB disk space. Use Scikit-learn, , Seaborn, OpenAI, NumPy, Pandas etc. to find the best results for pollutants that play an important role in increasing the AQI value. Python libraries (e.g. configuration files) are analyzed. The chart below shows AQI graphs for some pollutants that produce high AQI values. It is clear from Figure 7 that all pollutants rise and fall annually and their values ​​are not constant from year to year. PM2.5 and PM10 have an impact on winter, and air pollution in winter is more than in summer. After 2011, sulfur dioxide concentrations began to increase. The output can also be found at the BTX2 level. Almost all air pollutants, except CO, have seasonal variations. The data box and month divide all the data into different time periods. Figure 7 shows annual and monthly boxplots for different pollutants. Note that air pollution in India decreases between months of June and August. This may be due to the onset of monsoon in the Indian subcontinent this time. BTX levels decreased from March to April, increased slightly from September to May, and increased from October to December. The 2020 average is lower than previous years, indicating a decrease in air pollution in 2015. Development of machine learning AQI prediction model. Finally, performance of Air quality index prediction model is evaluated. Objective attitude, there are some missing values ​​in AQI, resulting

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

in disparity between groups. Most machine learning models ignore the problem of inconsistent data, which can lead to the prediction performance and poor classification. To overcome the inconsistent data problem, the GPT LLM model is dominated by the following parameters: 3.4-turbo-trainer, which is the main model of the open-text platform.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm used** | Kfoldcrossval mean score | R^2square | MSE | Data arrays |
| **Ridge regression** | 0.9345949 | 0.95 | 11.2214 | 60x12 |
| **Stepwise** | 0.952725727 | 0.97 | 7.59 | 60x12 |
| **polynomial** | 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 | 11.59303 | 60x12 |

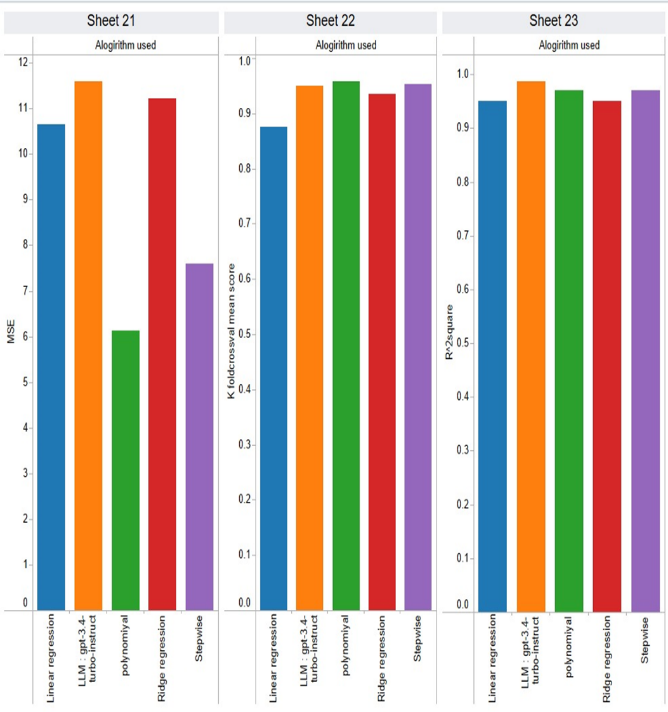
****

Fig.8 The various parameters of the algorithms

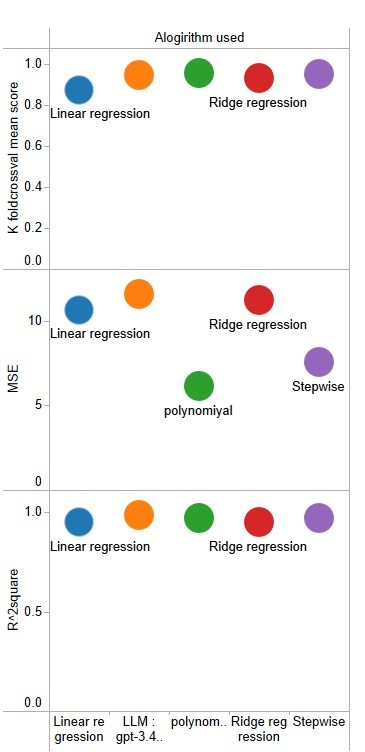
Correlations between various pollutants and ari quality indexes (AQI) are used and pollutants are used. This way, the algorithm synthesizes the new content for a small group rather than creating samples from existing products. It works by selecting a point from a limited class and counting the neighbors near the selected point. Newly created composite element will be added between the selected location and its neighbors. We use the AQI class asymmetry learning Python library to implement class asymmetric AQI. Currently, five popular ML models, linear and polynomial regression, ridge regression, stepwise regression and ChatGPT LLM model are used to predict the AQI level by AQI and AQI change method. Table 2 below shows results of recall, and MSE scores, accuracy, precision, for the ML model used during training. Precision refers to the proportion of samples that are relevant to a recall, while recall refers to proportion of sample retrieved. The fact is that the correct registration marks for all different systems are compared. Average precision and return scores.

Fig. 9 Variation analysis of pollutants

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:

**5.1 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.

**5. 2 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.

**5.3 Stepwise Regression:**

Low operating cost compared to LLM and polynomial regression.

Higher MSE and lower R2 value indicate poor performance.

**5.4 Ridge Regression:**

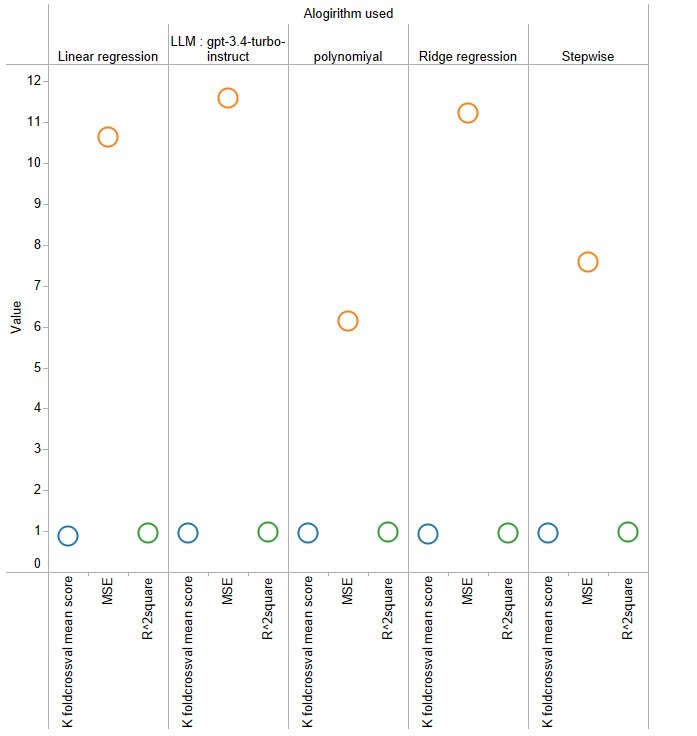
Records the worst performance of the comparison algorithms It has higher MSE and lower R2 values, which shows its limitations in comparison

The table above summarizes the performance of different ML models applied to different methods on the test set. It was observed that the LLM model showed improvement in almost all evaluation criteria compared to ML. The LLM model achieved the best value of R2 in both cases. The polynomial model performed the best in terms of error statistics and had the best fit in both test cases. These observations are highlighted in bold in Table 3.

## 6 Conclusion

In summary, our findings show that machine learning (ML) models, especially LLM (large linear mapping) models, outperform traditional methods in predicting the 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.



## 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.

Fig 10: Results of algorithms used and values

## References

1. 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)
2. 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>
3. 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)
4. 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>
5. 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>
6. 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)
7. 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>
8. 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>
9. 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)
10. 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>
11. 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>
12. 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>
13. 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>
14. 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>
15. 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)
16. 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)
17. 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
18. 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>
19. 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>
20. 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>
21. 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>
22. 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>
23. 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>
24. 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
25. 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>
26. 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>
27. 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 Compute. <https://doi.org/10.3390/bdcc2010005>