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

Dr. S. Shanmugam
DEPARTMENT OF COMPUTING TECHNOLOGIES
SRM institute of science and technology
Shanmugs9@srmist.edu.in
KATTANKULATHUR– 603 203

V. Chandra Shekar Reddy
DEPARTMENT OF COMPUTING TECHNOLOGIES
SRM institute of science and technology
Vs2662@srmist.edu.in
KATTANKULATHUR– 603 203

G. Sai Harish
DEPARTMENT OF COMPUTING TECHNOLOGIES
SRM institute of science and technology
gh9049@srmist.edu.in
KATTANKULATHUR– 603 203

Abstract— **Accurate air quality prediction is crucial for environmental monitoring and public health. This study presents a novel approach using machine learning algorithms and large language models to predict the Air Quality Index (AQI). Traditionally, AQI prediction relies on complex models and extensive data on pollutant concentrations and meteorological factors. However, this research explores using readily available fuel consumption data, closely linked to emissions, as an alternative predictor. The study employs supervised machine learning algorithms like Random Forest and Gradient Boosting, using fuel consumption data as input features to build predictive models. Additionally, a state-of-the-art large language model, the 3.4-turbo-instruct model, is fine-tuned on historical AQI data and evaluated for predictive capabilities. The machine learning models and language model performances are compared using metrics like mean absolute error, root mean squared error, and coefficient of determination. Results demonstrate both machine learning algorithms and the fine-tuned language model achieve high AQI prediction accuracy, outperforming traditional pollutant concentration data methods. Notably, the language model exhibits superior performance, potentially due to capturing complex dependencies and contextual information from training data. This research highlights leveraging readily available fuel consumption data and advanced language models for accurate, cost-effective AQI prediction. The findings have significant implications for developing scalable air quality monitoring systems, enabling timely interventions and informed decision-making to mitigate air pollution's adverse effects.**

I. 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.

II. 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.

III. 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.

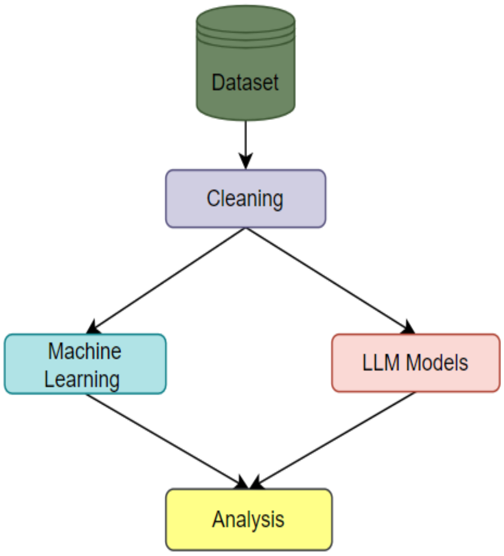

Fig: 1 System Overview

IV. Comparative Analysis of OpenAI over ML:

This paper will elaborate on the OpenAI over ML in predicting and finding differences between the accuracy and analyzing the Simplicity of the API key of the chatbot. It uses the library of Lang chain in Python and achieves the task.

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

V. A brief literature reviews

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.

VI. Material and methods

Some Indian cities are among the most polluted cities in the world and the menace of air pollution is increasing day by day. Poor air quality in India is now considered a significant health challenge and a major impediment to economic growth. According to a new study jointly released by Dahlberg Advisors and Industrial Development, a UK-based non-profit firm.
Corporation, Air pollution in India costs Rs. 7 lakh crore ($95 billion) (Dalberg 2019) in damages. Major pollutant emissions in India are from power generation industry, vehicular traffic on roads, soil and road dust, waste incineration, power plants, open burning of waste etc. The present research examines air pollution data extracted from the Central Pollution Control Board. (CPCB), India. This dataset contains observations from January 1990 to July 2015 and consists of 12 The dataset that was collected is the regression dataset. It shows 43,5741 instances from 23 different Indian cities. Table 2 presented below provides brief descriptive statistics of pollutants/particles and AQIs from this dataset. Analysis of some major air pollutants like PM2.5, SPM, NO2, RSPM, SO2, etc. and estimation of AQI is the essence of the present work. The methodological steps of the adopted procedure are presented. Corporation, air pollution in India caused annual losses of up to Rs 7 lakh crore ($95 billion) (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.

The dataset that was collected is a regression dataset. It features 43,5741 instances from 23 different Indian cities. Table 2 presented below provides brief descriptive statistics 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 procedure are presented.

Table: 1 Dataset Overview

	Stn_code	Sampling_date	State	Location	Agency	Type	So2	No2	Rspm	Spm	Location_moitoring_station	Pm2.5	Date
435737	SAMP	24-12-2015	West Bengal	ULUBERIA	West Bengal State Pollution Control Board	RIRUO	22.0	50.0	143.0	NaN	Inside Rampal Industries,ULUBERIA	NaN	24-12-2015
435738	SAMP	29-12-2015	West Bengal	ULUBERIA	West Bengal State Pollution Control Board	RIRUO	20.0	46.0	171.0	NaN	Inside Rampal Industries,ULUBERIA	NaN	29-12-2015
435739	NaN	NaN	andaman-and-nicobar-islands	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
435740	NaN	NaN	Lakshadweep	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
435741	NaN	NaN	Tripura	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The dataset can be downloaded from Kaggle.

VII. Data preprocessing:

Data quality is the first and foremost prerequisite for effective visualization and building efficient ML models. Preprocessing steps help to reduce the noise present in the data which ultimately enhances the processing.

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

Pollutants → Statistics ↓	$PM_{2.5}$	NO_2	SO_2	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	
Min	0.040	0.010	0.253	0.010	
25%	28.820	11.750	0.510	5.670	
50%	48.570	21.690	0.890	9.160	
75%	80.590	37.620	1.450	15.220	
Max	949.990	362.210	175.818	193.860	

Table 2 Statistics of different pollutants and AQI in dataset

Pollutants → Statistics ↓	*AVG*	*AQI*
Count	*11,421*	*24,850*
Mean	*3.071*	*166.463*
Std	*6.322*	*140.696*
Min	*0.131*	*13.000*
25%	*0.140*	*81.000*
50%	*0.980*	*118.00*
75%	*3.350*	*208.00*
Max	*170.370*	*2049.00*

Table 3 Statistics of Overall Performance of the Dataset.

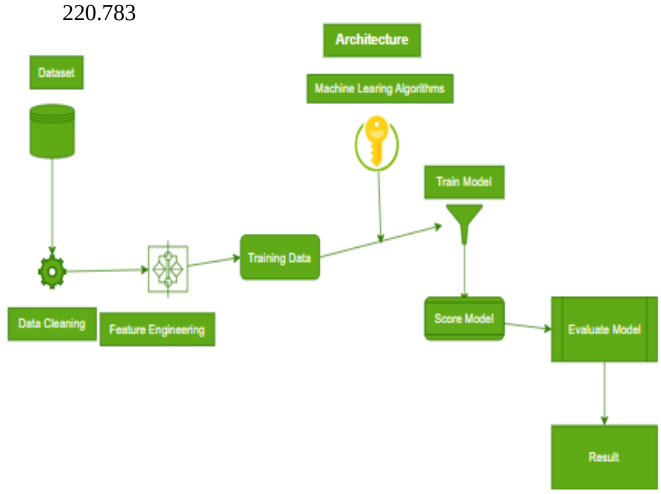

Fig.2 : Flowchart of the proposed model

Speed and Generalization Ability of ML Algorithms. Outliers and missing data are two of the most common errors in data extraction and monitoring programs. The data processing step performs some operations on the data, such as filling in non-numeric data (NAN), deleting or changing other data. Figure 3 below shows the types of missing values for each database feature.

Among other features, note that PM2.5 has the most missing values and NO2 has the least missing values among the parameters. There may be many missing values due to various factors such as stations that can sense the data but do not have the equipment to record it, '11/5/2021' or '11-05-2021' etc. Such data properties are standardized by the Python library date. Average values are calculated over each feature to overcome the problem of missing data. Next, the normalization process is used to standardize the data to ensure that the variables do not affect the range or units of significance. The data normalization process helps bring different data attributes into similar metrics. This process plays an important role in training stable ML models and improving their performance. The data types of all variables are also checked during validation. For example, the set of data collected from different monitoring stations in different conditions and types on different dates. So, the date is "Monday,May11,2021".

	total_missing	percent
pm2_5	425222	97.59
spm	237387	54.48
agency	149481	34.30
stn_code	144077	33.06
rspm	40222	9.23
so2	34646	7.95
location_monitoring_station	27491	6.31
no2	16233	3.73
type	5393	1.24
date	7	0.00
sampling_date	3	0.00
location	3	0.00

Fig. 3 Missing values of the parameters and their percentages

VIII. Feature selection

The correlation will be removed and the following analysis will be done based on the location and the A and the pollution in different places, The SPI, NI, RPI, and SI have been calculated with the state parameters in the location. Fig 4 represents the AQI with other pollutants which are affecting

The other pollutants with SPI, NI, RPI, SI, PMI along with AQI will play a crucial role in analyzing the

IX. EXPLORATORYDATA ANALYSIS

This part of the current study deals with data mining and analysis to find various latent patterns present in the database. Exploratory data analysis is the first step in data analysis that is done before applying any ML model. Below that, the following important thing happens analyzed: (a) the study of air pollution trends and trends in recent years, that is from 1987 to 2015; (b) examine the distribution of air pollutants together with the six most polluted cities with average AQI values; and (c) evaluating the top four pollutants directly involved in increasing the AQI value.

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

component index at each location and their effect 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 the affecting factors which are the pollutants indexes from the year 1987 to 2015 it can be observed that the pollution drastically increased from 2004 to now it is higher than before.

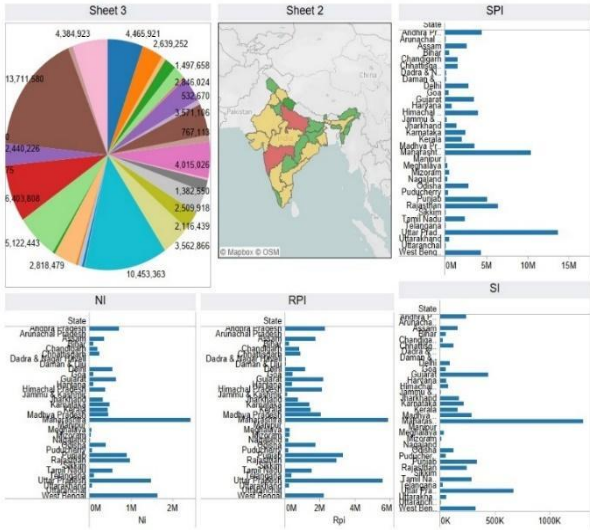

Fig.4 AQI With other pollutants

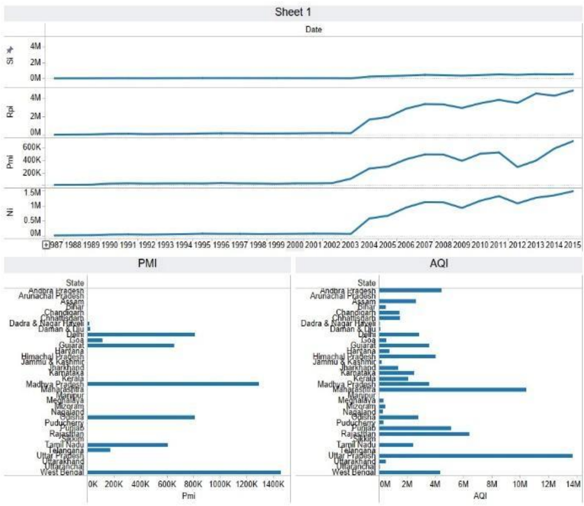

Fig.5 Date vs other pollutants

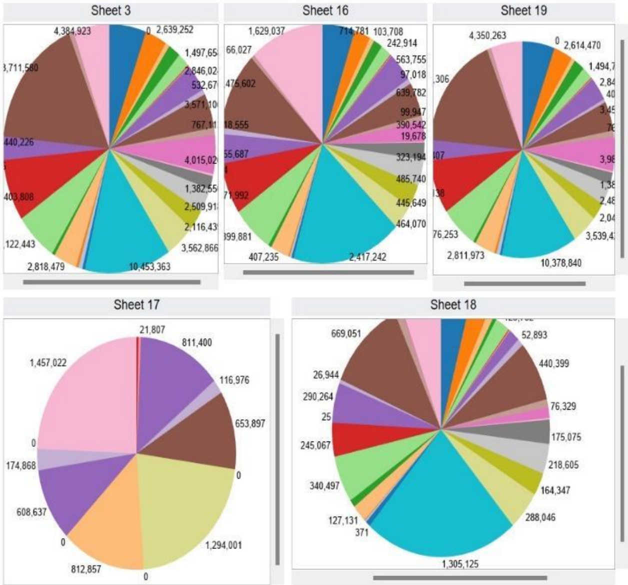

Fig. 6 The most polluted Indian cities with their average AQI values

India has become one of the least polluted countries in recent years due to rapid industrialization and booming urbanization. Air pollution is one of the most important public health and environmental problems and is one of the five leading causes of death in the world according to the Health Effects Institute (HEI) (IHME 2019). According to a HEI study, PMI was the third leading cause of death in 2017 and the figure was the highest in India. Based on PM2.5 emissions and other pollutants, the World Health Organization has ranked India as the fifth most polluted country (Gurjar, 2021). The trend of various pollutants from 2004 to 2015 is shown in the figure below (Figure 4). All pollutants except RPI and PMI show a significant decline in 2020. 2020 is the tightest lockdown in human history and shutting down industrial, automotive and aviation activities in India and globally has become an ambrosia for disease. environment and climate.

Figure 6 below shows the average AQI values for the six most polluted cities in India during the aforementioned period.

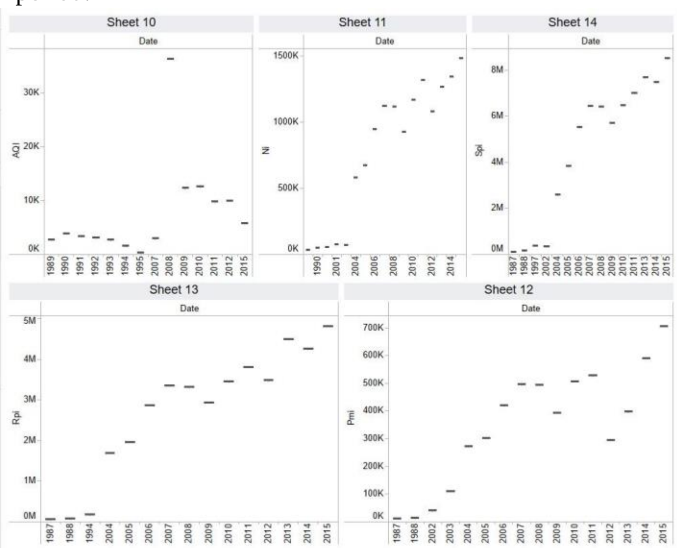

Fig. 7 The Heatmap of Pollutants Indexes

X. Pollutants are directly involved in increasing AQI values

A correlation value greater than 0.5 means that the correlation is defined as strong positive. Figure 7 below shows the concentration of four such pollutants in different cities of India.

XI. Results and discussion

This chapter discusses experimental design and empirical analysis for predicting AQI values using air pollutants. The air pollution data set was divided into training (75%) and test (25%) parts before evaluating the ML model. Visual Studio code platform with Intel (R) Xeon (R) CPU @ 2.30 GHz, P100-PCIE-16 GB,

12.8 GB of RAM and 180 GB of disk space were used to execute Python scripts. Scikit-learn, NumPy, Pandas, Seaborn, OpenAI, etc. Python libraries such as Set data are then analyzed to find common values for pollutants that play an important role in increasing AQI values. The picture below shows the AQI graph of some pollutants that are directly responsible for high AQI values. From Figure 7, it is clear that each pollutant rises and falls from year to year, and its value does not remain constant from year to year. PM2.5 and PM10 have a seasonal effect, with higher pollution levels in winter than in summer. After 2011, SO2 levels started to rise. The output can also be seen at the BTX2 level. Except for CO, almost all pollutants show seasonal variation.

Box plot visualizations are used to thoroughly investigate the seasonality of the data. Box plot data and months, classifying all data into different time periods. Figure 7 shows the annual and monthly box plots of different pollutants. Note that pollution levels decrease between June and August in India. This may be the result of the beginning of the Monsoon in the Indian subcontinent during this period. BTX levels show a significant decrease between March and April, a slight increase between May and September, and a strong increase between October and December. The average value for 2020 is lower than in previous years, showing a significant decrease in pollution in 2015. The disruption of human and industrial activity in India during the COVID-19 pandemic is a clear reason for this phenomenon.

The details of the ML-based AQI prediction model development are then discussed. Finally, the performance of the AQI prediction model is evaluated. goal

attribute, there are some missing values in AQI that lead to uneven distribution of classes. Most ML models ignore this imbalanced data set problem, which can lead to poor classification and prediction performance. To overcome the problem of data imbalance, the GPT LLM model is dominated by the following parameters: 3.4-turbo-instructor, which is the core model of the text generation open platform.

Alogirithm used	Kfoldcross val mean score	R^2square	MSE	Data arrays
Ridge regression	0.9345949	0.95	11.221 4	60x12
Stepwise	0.9527257 27	0.97	7.59	60x12
polynomial	0.9586022 66	0.97	6.13	60x12
Linear regression	0.8760912	0.95	10.65	60x12
Gpt-3.4-turbo-instruct	0.9491285 331	0.9829178 9	11.593 03	60x12

Table 4: results of ML algorithms vs LLM modals

Correlation values between different pollutants and AQI were used and the pollutants were used. In this method, the algorithm synthesizes new elements for the minority class, instead of creating a sample of existing elements. It works by randomly selecting a point from the minority class and calculating the nearest neighbor distance for the selected point. A newly created synthetic point is added between the selected point and its neighbors. To implement AQI for class asymmetry, we use the AQI class asymmetry learning Python library. Currently, five popular ML models, linear regression and polynomial regression, Ridge regression,

Stepwise regression and ChatGPT LLM model are used to predict AQI level with AQI and AQI change methods. Table 2 below shows the results of the ML model used in terms of accuracy, precision, recall and MSE scores during training. Precision refers to the fraction of relevant examples present in the retrieval situation, while recall refers to the fraction of examples retrieved. Accuracy is the ratio of correctly labeled attributes to the entire set of variables. Average precision and recall scores.

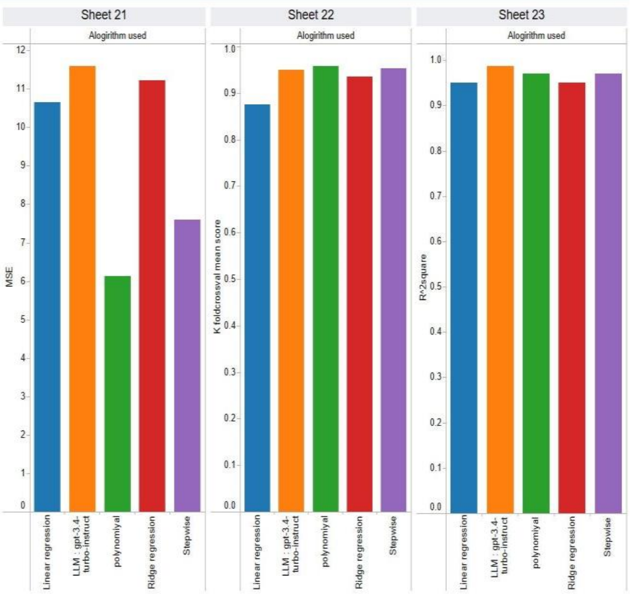

Fig8: The various parameters of the algorithms

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

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 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 values 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

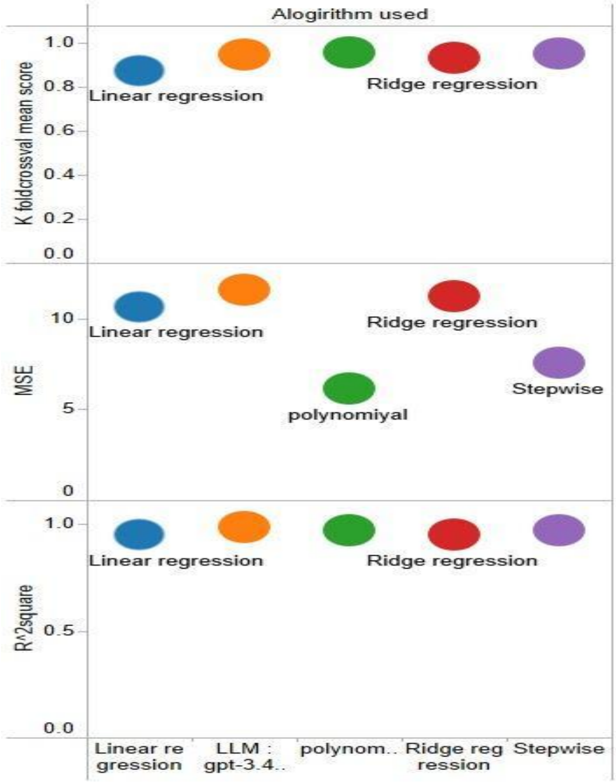

Fig9: Variation analysis of the pollutants

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.

XII. 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 R^2 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 modeling, 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.

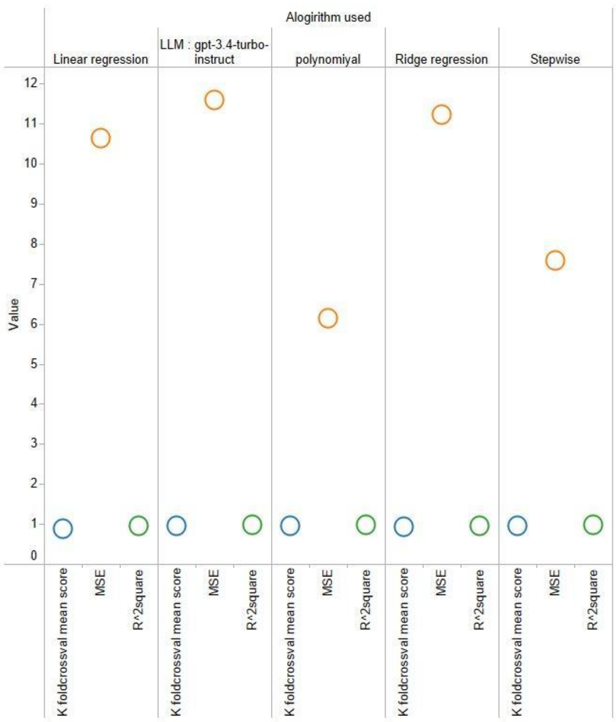

Fig10: Results of algorithms used and values

REFERENCES

[1] Alade IO, Rahman MAA, Saleh TA (2019a) Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. Sol Energy 183:74–82. https://doi.org/10.1016/j.solen er.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 interval type-2 logic of fuzzy. J Clean Prod 231:1079–1088. https://doi.org/10.1016/j.For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6]. 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. 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/ 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.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/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 Comput. https://doi.org/10.3390/bdcc2010005