Inspira-Journal of Commerce, Economics & Computer Science (JCECS)
ISSN : 2395-7069(Print), Impact Factor: (Cosmos 7.650 & I2OR 5.120), Volume 12, No. 01, January-March, 2026, pp. 103-108

Monitoring Statewise Annual Average PM Value for Understanding Air Pollution Along with Determination of Trend by Artificial Intelligence and Machine Learning

Sumana Chatterjee*

Ph.D. Scholar (Computer Science), Nirwan University Jaipur, Rajasthan, India.

*Corresponding Author: sumana.spssaha.chatterjee@gmail.com

Citation: *Chatterjee, S. (2026). Monitoring Statewise Annual Average PM Value for Understanding Air Pollution Along with Determination of Trend by Artificial Intelligence and Machine Learning. Inspira-Journal of Commerce, Economics & Computer Science (JCECS), 12(01), 103–108. https://doi.org/10.62823/JCECS/12.01.8552*

Abstract

This paper is based on data analysis by python programming language code, data executed on google collaborator platform with PM, i.e. particulate matter data of NAMP, a nationwide air quality monitoring programme, data available in site of CPCB, central pollution control board. The objective of data analysis was to monitor state wise change of annual average value of particulate matter, PM, for past few years, years as available in the site of CPCB and obtained accordingly. The year wise NAMP data, for 2013 to 2023 as available there were collected and subjected to analysis for understanding year wise status of air pollution influenced by PM value. The sources of PM or particulate matter are vehicular emissions, coal based power plants, construction activities, road dust, biomass and garbage burning, industrial processes etc. So to monitor the PM level obviously matters a lot as presence of PM with higher level can impact on health causing serious respiratory diseases like asthma, bronchitis, also heart disease even stoke. Different types of PM are there depending on diameter such as PM10, PM2.5, PM1, measurement units micrograms per cubic metre. Depending on availability either PM10 or PM2.5 were collected and subjected to analysis defined as PM value. Visualization of yearly change as well as trend of PM level for the year 2026 for each state had been determined with the help of artificial intelligence and machine learning. Line plots, bar plots were obtained to visualize year wise changes and neural network model was used to find current year trend.

Keywords: *Air Pollution, PM, NAMP, CPCB, Python, Google Collaborator, Line Plot, Bar Plot, Neural Network.*

Introduction

Air pollution is a curse of urbanization and industrialization. Dust caused by urban construction like construction of houses, flats, high rise buildings etc. , smoke caused by continuous emissions from vehicles ,emissions from power plants ,coal based plants ,biomass and waste burning are main sources of air pollution . There are also other natural sources like dust storms and forest fires . All these types of sources of air pollution continuously increase level of particulate matter or PM . Different types of PM are there depending on its diameter such as PM_{10} , $PM_{2.5}$ and PM_1 depending on the size of the diameter respectively with ≤10 μm , ≤2.5 μm and ≤1 μm . As small as the diameter of it is, it is much more dangerous as then it can penetrate much more easily into human respiratory system .Presence of PM substances in air with level higher than the safety level can lead to higher risk on human health. The other factors which can cause air pollution are SO_2, NO_2, CO etc. but among all these, PM causes most to invite several respiratory and lungs diseases and also other critical diseases related to heart, skin,

damaging other various human organs silently . For this reason it is necessary to monitor the level of PM value year wise and state wise with past data from NAMP , 'national ambient air quality monitoring programme' implemented by central pollution control board in collaboration with state pollution control board (SPCB) and pollution control committees (PCC) . Year wise NAMP data obtained in pdf format combined with all air pollutants such as SO_2, NO_2, PM_{10}, $PM_{2.5}$ etc. , among which annual average PM value was extracted for analysis for each state in India to understand status of PM level influencing air pollution . For the year 2013 , not the $PM_{2.5}$ but only PM_{10} value was available and for other years , $PM_{2.5}$ value was available for extraction from the csv files . Thus obtained values for particulate matters , observation taken for some days , year wise, state wise ,location wise ,annual average particulate matter value defined as PM data common for all years ,were subjected to analysis.

Literature Review

The research paper is based on analysis of pollution data for some years starting from 2013 to 2023 for visualization of status of presence of particulate matter in air . All the related research papers relevant with this appropriate subjects have been studied and reviewed the literature and the current study had been done almost in similar manner . The data collected was from the online site of CPCB and NAMP annual monitoring pollution data . Different data sources were available online for collection of air pollution data but to avail continuous data starting from some past years , the data of NAMP was chosen for analysis. For visualization of status of particulate matter PM value ,similar as reviewed literature of other research papers based on this subject that is study of status of air pollution from past to recent years ,machine learning technique had been applied for visualization with bar plots where as the predicted value for 2026 was found by the analysis with neural network and the analysis by Time-Series CV + Lag Features . The size of data collected, as available on the online platform, was not sufficiently enough for prediction by neural network as successful model prediction is generally used with large historical data for neural network model to obtain higher accuracy . So another type of analysis by time series CV and lag features which can give better accuracy with data with small size was executed. But overall the average model accuracy for neural network was higher and RMSE much more less than ridge regression model . The accuracy score for some states suffered due to incomplete data for some years ,the problem with such non-availability could not be resolved .The sparsity of data also affected the robust model with Time-Series CV + Lag Features. However the comparison between RMSE score of two types of analysis was overviewed to get insights regarding predicted value for current year.

Research Gap

The paper based on artificial intelligence and machine learning to visualize the status of PM level from past years for different states of India, data used from annual monitoring programme of central pollution control board . Instead of analysis of hourly , daily air quality index value , the annual average particulate matter value for different states over India starting from past years were subjected to analysis to get insight of yearly annual trend and probable annual trend of 2026. All the execution and python analysis was subjected to the data collected from online NAMP data as available there. According to the data availability a thorough analysis had been done to obtain the state wise annual average PM value of past years for Indian state as well as predicted value for current year for the same.

Research Questions /Hypothesis

The research problem was to study the level of existence of particulate matter in the air , analysis based on year wise data for all the states of India . This analysis could make us understand the status of PM level from past years, 2013 to 2023 and based on this analysis with the help of neural network , secondly it was to determine the probable trend of the level of presence of particulate matter in the air for the current year 2026, which also was performed with the help of analysis by artificial intelligence and machine learning based on obtained data as per availability. The research question was to monitor whether and where ,region wise, state wise, there is increasing or decreasing trend in recent future and what was the trend for past years . Accordingly by analysis this could be obtained for Indian states ,obtained individual results.

Methods

This paper is based on analysis with air pollution data , collected from the online site of CPCB ,central pollution control board and from there available data from NAMP i.e. National Air Monitoring Programme for the years 2013 to 2023 as available there to download . Analysis by python programming

language was done with the downloaded data to visualize the status of air pollution from available past years and to predict status of air pollution for current year .For visualization of status of air pollution in past available years the technique of machine learning was applied .For prediction of probable status of air quality i.e. existence of particulate matters influencing pollution ,the process of analysis with neural network was used . Moreover the analysis was also done by Time series CV and Lag method with training Ridge Regression model as this model is used for better accuracy with small size of data set . But ultimately as result obtained , the accuracy score with neural network model could be obtained as higher value than that of ridge regression model . Annual average PM value for some states suffered due to the gap of downloaded data. The first phase was to import python libraries necessary for execution on google collaborator platform . Then after this uploaded the csv file with data of particulate matter PM from 2013 to 2023 . The air pollution data of NAMP observation was mixed with all pollution factors such as SO_2, NO_2, PM_{10}, $PM_{2.5}$ etc. among these the particulate matter PM values were extracted for analysis . For 2013 , PM_{10} was available and chosen as factor of analysis representing particulate matter in air and for other years from 2014 onwards , $PM_{2.5}$ was chosen where both of PM_{10} and $PM_{2.5}$ was available in the particular year's data. From the available data set for years , filtered the column with value of particulate matters and extracted as different csv files for analysis. Another file combining all these individual data frames represented a single comprehensive data frame .The comprehensive data frame as well as individual data files all were necessary for several types of analysis to get various insights. Feature engineering and data cleaning process was done , renaming the target column with same name for each csv files for each year ,for the purpose of analysis. Dropped the column with the null or missing values. In some of data files the name and spelling of the name of same state were used in different manners and for the purpose of analysis, there the state name had been standardized for the purpose of proper result from analysis. The combined csv data file was used to visualize line plot and trend of particulate matter in air ,state wise trend from past years along with probable predicted value for current year. For visualization, the necessary python libraries as used were 'matplotlib' and 'seaborn'. One by one ,status of existence of particulate matter PM value from past to recent year along with predicted value for 2026 was subjected to visualization for each state separately. Necessary insights could be overviewed from these visualizations. For the case of Andaman and Nicobar (UT) , dropping rows with null values resulting into an empty data frame for this state which was the reason of having very insufficient observational data for this state which was managed accordingly. The prediction for 2026 suggested a continuing low PM value for Himachal Pradesh and all the information as obtained from the insights of python analysis may add insights for environmental policy making and health planning . Sparsity of data was a problem of some of the states which could not be overcome due to non-availability of data. Sequential , tensor flow based neural network model was used for analysis and prediction. Activation function used was 'Relu' with optimiser as 'Adam'. Prediction of a high PM value for 2026 ,significantly higher for Orissa was obtained. For Tamilnadu , prediction for 2026, suggested a relatively low PM value indicated status of improving air quality .For Mizoram, suggested a very low value of PM .The predicted value of 2026 suggested a relatively high PM value for Bihar and moderate value for Uttarakhand. Similarly trend for other states could be obtained. The north west states had higher PM value for past years , bar column for visualization as obtained. Delhi , Jharkhand, Rajasthan and north west states had higher values of annual average PM values. The southern and north eastern states had lower PM value as prediction obtained. Other process of analysis was also followed which was time series CV and lag features by training ridge regression model. Generally this model works well with small size of data but the accuracy score as obtained here was much lower for most of the states than that of neural network model. Lagged PM values (lag1,lag2,lag3) for 2024,2025 and 2026 were incorporated with a trend feature for each state. Prediction was done by iterative forecasting . The average RMSE score for ridge regression model as obtained was much much higher than that for neural network model. State wise RMSE score was also obtained for neural network model to visualize the model performance. Individually state wise trend for past years from 2013 to 2023 along with predicted value for 2026 for each state was visualized by line plot.

Significance of the Study

The status of existence of particulate matter influencing air pollution could be visualized by the study . As per Indian standard (CPCB-NAAQS) ,the standard safety level for $PM_{2.5}$ annual mean≤40 $\mu g/m^3$ and for PM_{10} annual mean ≤60 $\mu g/m^3$. But some small amount of $PM_{2.5}$ is also not safe as this tiny particulate matter can penetrate into lungs and initiate cause of various respiratory diseases

.Moreover existence of particulate matter in air can damage heart ,skin etc. of human beings ,so the monitoring of status of average annual value of particulate matter for past years was very essential for environmental policy making and health planning. From this study ,the change and status of PM value ,trend of the same could be visualized . Moreover the predicted PM value for 2026 i.e. the probable annual average value of particulate matter for recent future ,increasing or decreasing trend could also be visualized .

Timeline

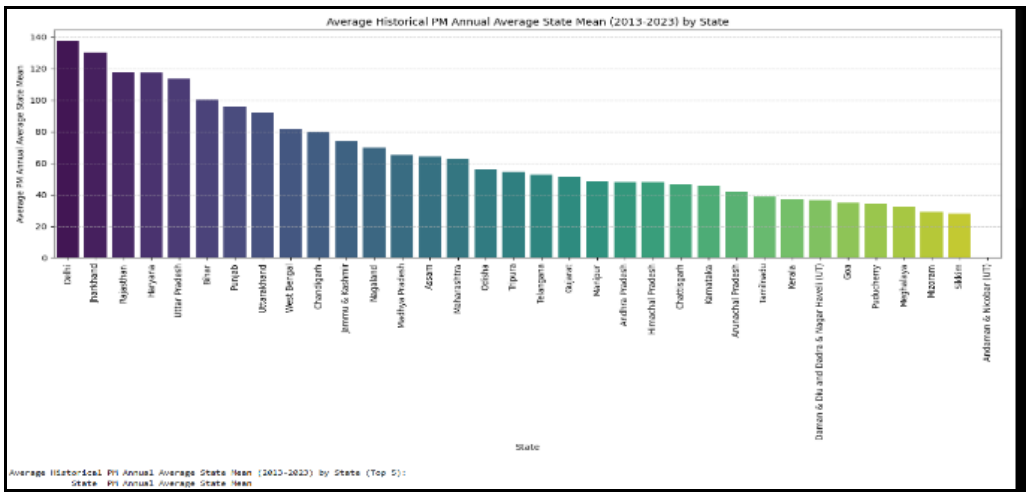
The collection of data from online site CPCB , central pollution control board and the annual monitoring data NAMP for past years as available from 2013 to 2023 were collected for analysis. Then sorting and filtering of the particular column , year wise extraction of annual observational particulate matter data were subjected to python analysis to get insights to know the status of trend of air pollution. The extraction of data , feature engineering ,execution as well as compatibility, model fitment ,analysis , all had to be done as applicable therewith.

Conclusion and Future Work

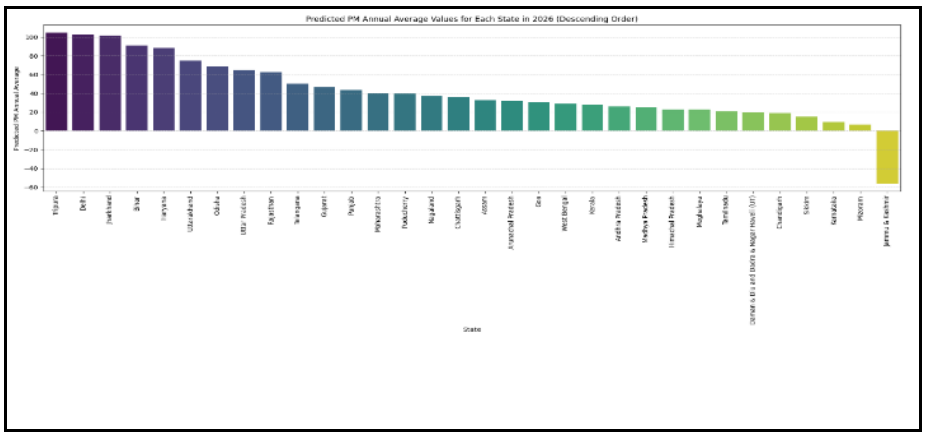
The whole analysis and result of model determined , was based on available data . Due to the sparsity of data ,the analysis for some states was affected . For this reason another time series ridge regression model was applied which generally is applied for small size data set . But as result obtained , the average RMSE with minimum value for the neural network indicated the higher success rate. The historical data as available from other online data collection platforms to collect air pollution data , availability as checked therein ,data was to collect from multiple platforms , for different years ,not from any single source . So to avoid this problem with multiple sources of online platforms , the online NAMP data was collected as available there but the analysis faced somehow sparsity issue. In future analysis , visualization of trend may be done for any single state of India , data collection to be done accordingly for historical data for single state depending on availability. As the presence of particulate matter causes a lot for the reason $PM_{2.5}$ is more dangerous ,penetrates deep into lungs ,enters bloodstream ,triggers heart disease ,stroke, diabetes ,cognitive decline ,even can cause premature death also ,so analysis for understanding of trend ,with other state or region having sufficient observational data to get valuable insights accordingly. In future analysis with air pollution data collected from other sources, may be subjected to analysis to get output as well as predicted value. Moreover study with influence of particulate matter on meteorological parameters may be done also in future to get insights from that. If possible ,it will be tried to overcome the problem with data sparsity as faced with this research study .

Some Screenshots of Analysis

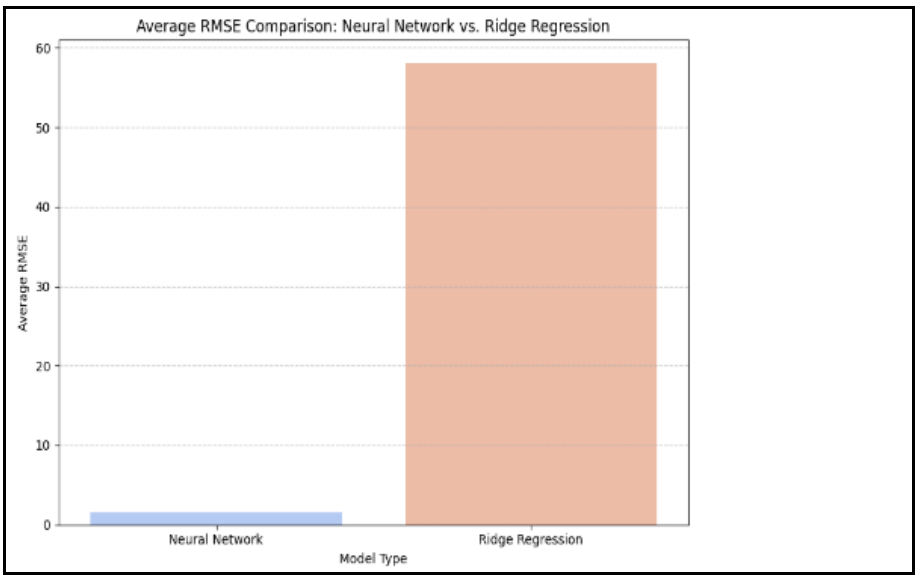
(State Wise Annual Average Of PM Value2013-2023) (State Wise Predicted Annual Average PM Value 2026)

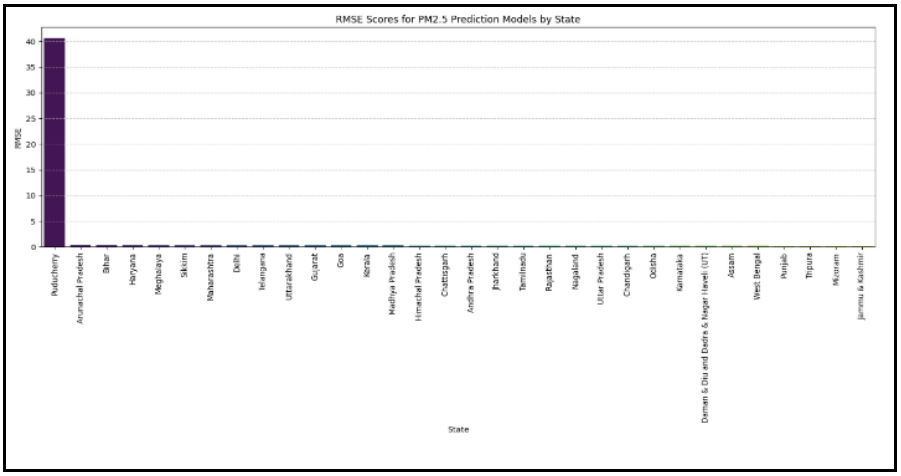


Average RMSE Comparison, Neural Net Work and Ridge Regression



State Wise RMSE Scores for Predicted Value 2026





References

1. Chadalavada, S., Faust, O., Salvi, M., Seoni, S., Raj, N., Raghavendra, U., Gudigar, A., Barua, P. D., Molinari, F., & Acharya, R. (2025). Application of artificial intelligence in air pollution monitoring and forecasting: A systematic review. Environmental Modelling & Software, 185, 106312. [1] https://doi.org/10.1016/j.envsoft.2024.106312
2. Zhou, S., Wang, W., Zhu, L., Qiao, Q., & Kang, Y. (2024). Deep-learning architecture for PM2.5 concentration prediction: A review [1] [2] [3] [4] [5] [6] [7]. Environmental Science and Ecotechnology, 21, 100400. https://doi.org/10.1016/j.ese.2024.100400
3. Karmoude, M., Munhungewarwa, B., & Chiraira, I. (2025). Machine learning for air quality prediction and data analysis: Review on recent *advancements, challenges, and outlooks*. *Science of the Total Environment, 1002*, Article 180593.

   https://doi.org/10.1016/j.scitotenv.2025.180593
4. Latoń, D., Grela, J., Ożadowicz, A., & Wisniewski, Ł. (2025). *Artificial intelligence and machine learning approaches for indoor air quality prediction: A comprehensive review of methods and applications*. **Energies, 18**(19), 5194. https://doi.org/10.3390/en18195194
5. Chen, Y., Wu, Y., Zhang, S., Yuan, K., Huang, J., Shi, D., & Hu, S. (2025). *Regional PM2.5 prediction with hybrid directed graph neural networks and spatio-temporal fusion of meteorological factors*. **Environmental Pollution, 366**, 125404.

   https://doi.org/10.1016/j.envpol.2024.125404.

❑❑❑