JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A UNIQUE ALGORITHM FOR CONVENTIONAL STATISTICAL MODELS WITH DEEP LEARNING FOR PREDICTION OF AIR QUALITY

Jaladi Naga Sireesha, Dr. Jithendra P R Nayak, Dr. Rajaram
Asst. Professor, Associate Professor, Associate Professor
Computer Science & Engineering
Malla Reddy Engineering College for Women, Hyderabad, India

Abstract: Air pollution as well as poor ventilation is given by the quick rise of urbanization and industries. It is a "silent public health emergency" due to its damaging consequences on equally people and the environment. perfect air pollution forecast is essential for stakeholders for taking the essential action in order to address this worldwide concern. Compared to other methods, deep learning-based models for prediction have shown promise in the past few years for precise and effective air excellence forecasting. In order to anticipate five air pollutants, including nitrogen dioxide (NO 2), ozone (O3), sulfur dioxide (SO 2), and long short term memory (LSTM) and particulate matter (PM 2.5 as well as PM 10) we conducted a comparison examination of several deep learning-powered single-step forecasting techniques in this research. We utilized a publically accessible dataset obtained via Kaggle called "Air Quality Data in India (2015 – 2020)" for our empirical analysis. It calculates the amount of air pollution present. Three indicators of performance are used to assess the effectiveness of forecasting models: R-squared (R2), mean absolute error (MAE), and root mean square error (RMSE) by a score of 0.59, the outcome demonstrates that machine learning models consistently produced the lowest RMSE when weighed against statistical models. Furthermore, it is discovered that the deep learning model has the greatest R2 score of 0.856.

***IndexTerms* - Deep learning models mean absolute error (MAE), root mean square error (RSME), air pollution.**

Introduction

In recent times, air pollution has emerged as a important worldwide issue. Human well-being and health are directly impacted by air pollution, in addition to its negative belongings on the ecosystem. It has been noted that illnesses related to breathing, cognitive decline, cardiovascular disorders, and cancer are among the morbidities and deaths that are increased by air pollution. More than 3 million deaths are attributed to air pollution annually, primarily in low- and middle-income nations. Furthermore, by enhancing air quality along with other aspects, the United Nations (UN) has established sustainability goals (SDG) 3, 7, and 11, which set targets for 2030 to lower mortality, disease, and the negative environmental impact in cities. Comparably, in the UK, where the government is set a target to decrease by 35% of air pollution by 2040. decomposing air quality is caused by a number of variables, including producing goods, industrial pollutants, dust, coal usage, and emissions from land, air, and sea traffic. The release of toxic materials as well as gases into the atmosphere is known as air pollution, and it is a major health risk to people as well as living things. Pollutants are these dangerous substances (solids, liquids, or gases). When some pollutants, such as PM2.5 (particulates matter), generate in greater quantities than typical, it degrades the surroundings and has detrimental impacts on human health.

Educating the public and raising consciousness of this issue calls for multidisciplinary approaches including experts and other stakeholders. As part of their responsibility, local governments have established several air quality check facilities across the nation to track the amount of air contaminants. Pollutant predictions can be made using information gathered from these surveillance sites. Air quality forecasting is essential for both reducing air pollution and identifying areas that require remediation to lessen its effects. However, precisely predicting air quality is a challenging endeavor that depends on the modeling methodologies and accessible data. The primary findings of this research include Educating the public and raising consciousness of this issue calls for multidisciplinary approaches including experts and other stakeholders. As part of their responsibility, local governments have established several air quality monitoring facilities across the nation to track the amount of air contaminants. Pollutant predictions can be made using information gathered from these surveillance sites. Air quality forecasting is vital to both minimize air pollution and identify areas that require remediation to lessen its impacts. Yet, precisely predicting air quality is a challenging undertaking that depends on the modeling methodologies as well as information that are accessible. The primary findings of this research include

Urban air pollution has a negative impact on ecosystems, financial markets, and public health, making it major global health including safety concern. Urban regions now contain higher concentrations of pollutants similar to particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), and ozone (O3) due to the fast growth in urbanization and industrialization that has occurred in recent decades. Amongst other sources, the main producers of these pollutants include power plants, automobiles, industries, and domestic heating.

Because urban air pollution is complicated and dynamic, it is essential to accurately estimate pollutant concentrations in order to regulate air quality, formulate policies, and safeguard public health. By utilizing statistical correlations among pollutant concentrations and different meteorological, geographic, and human characteristics, conventional statistical approaches have become widely employed for the prediction of air pollutants. fortunately is increasing interest in examining the effectiveness of deep learning algorithms for air pollutant prediction due to the development of methods for deep learning and the accessibility of large-scale pollutants data.

Literature Survey

Particulate matter (PM) as well as gases make up the heterogeneous mixture that is known as air pollution. PM in today's urban environment is mostly caused by burning fossil fuels, and each component can range in size between a few nanometers to ten micrometers. The source of pollution and its chemical make up along with ambient levels, may have an impact on the biological toxicity and ensuing health impacts. However, research conducted worldwide has repeatedly demonstrated that exposition to PM, both short- and long-term, is linked to a variety of cardiovascular disorders, such as myocardial infarction as well as infarctions that, cardiac failure, arrhythmias, strokes, and an elevated risk of cardiovascular death.

Daily fatalities appear to be strongly linked to temporary publicity to air pollution; yet, the majority of the evidence for this relationship is derived from studies carried out in large cities, as well as little has been learned about the degree of spatial heterogeneity within the consequences throughout areas that include either urban while non-urban settings.

Our objectives were to examine the potential short-term correlation among air pollution and cause-specific death on a daily basis in the province of Stockholm, as well as to determine if this correlation extends beyond the urban area. For the period of 2005–2016, we employed a spatiotemporal random forest approach to forecast daily amounts of ozone (O3), nitrogen dioxide (NO2), and tiny and inhalable particulates (PM2.5 and PM10) over Sweden using a spatial resolution of 1 km. For market research, we gathered daily mortality data for every small area (SAMS) of the Stockholm County, that we corresponding daily exposure to air pollutants and air temperature.

"The largest probable threat to the general population throughout the UK" is air pollution, according to the government. The environment as well as economic expansion are also impacted by air pollution, in addition to human health. International agreements, EU legislation, national and devolved legislation, and the cross-border nature of air pollution have all impacted attempts to control and improve air quality across the United Kingdom.

Air pollution is one well-known environmental health danger. We know what we're striving for when a plume emerges from a smokestack, emissions spread over a major thoroughfare, or a brown cloud appears throughout a metropolis. Even while some air pollution is invisible, it can still be identified by its overpowering smell.

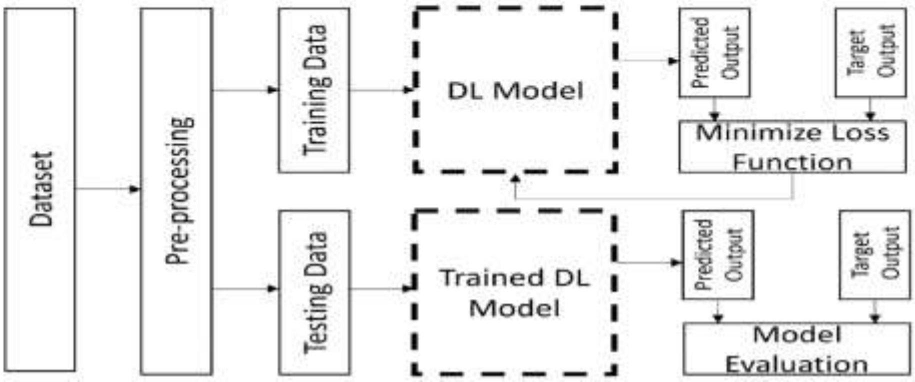
It puts one's health and fortune at significant danger of the world. All forms of air pollution are thought to be responsible for about 6.5 million deaths globally, a number that has increased over the past 20 years.

Existing System

Statistical Methods: Studies on environmental forecasting indicate that these have been used in some form throughout the 1970s. They create statistical models for forecasting future concentrations by utilizing past data on air contaminants, weather, among other variables.

Dispersion modeling is a method that is being applied since the middle of the 20th century.

It makes predictions about how contaminants will spread throughout the climate using computer simulations that take into account geography, wind direction, and speed.



Proposed System

Because deep learning models effectively several important advantages beyond traditional approaches, they have revolutionized the field about air quality prediction. This is a comparison between deep learning and the current models.

A data flow chart is a visual aid worn to demonstrate how information moves across a classification or operation. Information as regards every entity's components and results in addition to the process overall are also present by the DFD. Here are no loops or decision rules in an illustration of data flow, so there is no way to regulate flow.

Model Training and Testing

Information regarding the single-step prediction models' evaluation, hyperparameter optimizing, training, and data processing is provided in this part of the manual.

Data Pre-Processing

In overall, any dataset could have erroneous values therefore outliers, and it might require to be standardized to meet the needs of the forecasting model. Excessive or unusual values that deviate from the average in a dataset are called outliers, and their existence can have an impact on the pattern of distribution of data as a whole. To enhance the accuracy of the forecasting model, anomalies must be eliminated. Similarly, prior to modeling, any deficient or periodic pattern of values in the dataset—known as invalid values—must be eliminated or substituted with some estimated values. Lastly, the dataset is normalized by rescaling the data to fit inside the specified range. The normalization process can accelerate prediction models' converging and help them operate effectively on smaller-scale data. This interquartile range approach (IQR) is employed to preprocess the dataset presented in this article so as to identify outliers and eliminate erroneous values. We organized the data by month, day, and hour during pre-processing in order to substitute the missing values. Next, the values that are missing are substituted by averaging the concentration values that are accessible on the same month, day, and hour over every year in the data set. A wider range of values can be accommodated for the information that is lacking using this method.

For prediction models, weather information is regarded as an additional input element. In addition to the characteristics found in the dataset, the researchers additionally produced a lag feature by dividing the date and time index into two halves and using the percentage value of the pollutant predicted from the previous hour, day, and month to generate additional features. As a result, in order to estimate every pollutant, we took into account a variety of factors, including the concentration from the hour before, meteorological data (temperature, wind direction, and speed), and temporal data (day, month, and hour) has expected outcome of the following hour. Fig. 1 illustrates the pre-processing approach for the data set. Utilizing the Min-Max standardization defined by, wherein xmin as well as xmax are the lowest and highest points of data x, the input characteristics are normalized. A small amount of NO_2 is displayed in Fig. 2 both prior to and following pre-processing data.

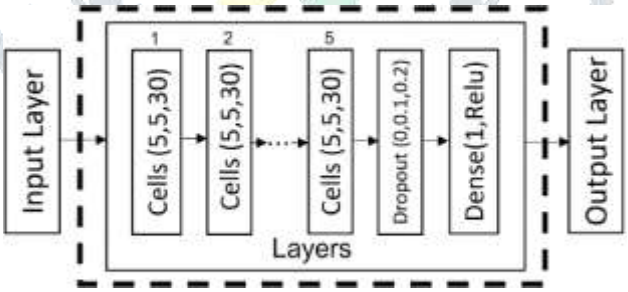


Fig.2. Architecture of deep learning model

Dropout rates is also taken into consideration as an optimization parameter in the dropping layer, and which arbitrarily removes a certain amount of cells to manage overfitting. Finally, a dense layer that is entirely coupled and has a Relu activation function is employed. The Hyperband technique [31], which adjusts the number of modules in a layer, the quantity of layers employed, and the pace of learning of the model, is employed to tune the hyperparameters while the Adam optimizer is employed for training. In order to shorten the training period, the Hyperband technique minimizes the validation loss throughout model training in order to optimize the hyperparameters. Table 3 provides an overview of the deep learning models' design characteristics together with a description of their parameters. The LSTM model employed the lowest possible number of layers—two for SO_2 and a maximum of four for NO_2—after hyperparameter optimization. In contrast, the GRU model employed a maximum of four layers for PM10 as well as a minimum of one layer for PM2.5. Additionally, for the pollutants under consideration, the LSTM and GRU methods employed an optimized cell count within the ranges of [35 70] as well as [20 70], correspondingly. Furthermore, optimized dropout rates having a value of 0 or 0.1 for each pollutant were applied in both DL models. Table 4 provides the details of the ideal set of hyper parameters for each pollutant, and these variables can also be utilized to assess the computing effort and sophistication.

LSTM (Long Short-Term Memory)

Recurrent neural networks (RNNs) are networks that contain loops to preserve knowledge. RNNs are able to use the previous information whenever there is a tiny gap among the associated data and the location here it is needed. Regretfully, RNNs lose their capability to be taught to connect the data as the gap widens.

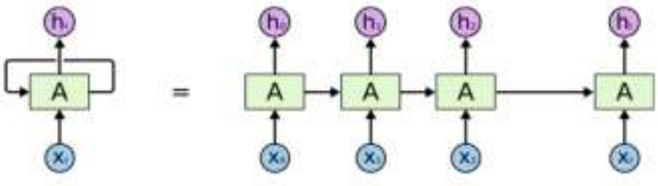


Figure 3 Recurrent Neural Network

LSTMs are a unique type of RNN that can adjust to extended periods of time. Long-term information retention is their standard mode of operation. Similar to RNNs, LSTMs also feature a chain-like framework, however the recurrent module features a different structure. Instead of an individual neural network, there's are actually four layers that each work in a different way. The cell state is the path to LSTM. The condition of the cell is comparable to that of a continuous belt. It has a few small linear connections but otherwise flows directly down the chain. Knowledge about the cell state can be easily accessed and is meticulously regulated by structures known as gates.

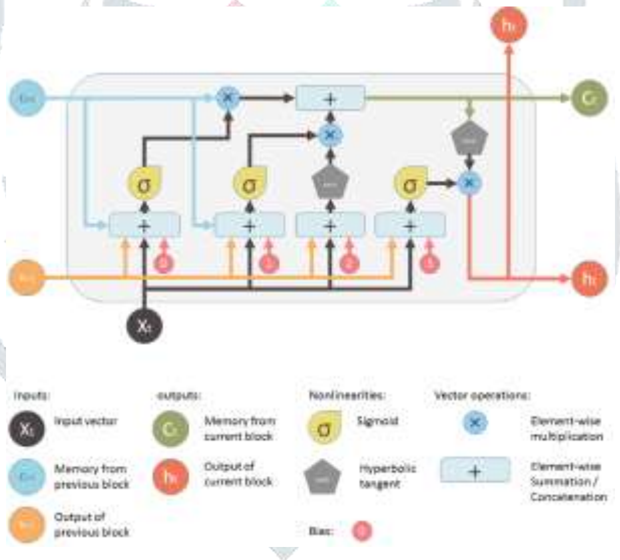


Figure.4. Basic LSTM Memory Block

Considering there might be lags of unclear length among important happening in a time series, LSTM networks are appropriate for categorization, processing, and forecast according to time series data. These were developed to handle the issues with vanishing as well as bursting gradients that arise throughout the training of conventional RNNs. The logistic function is often the LSTM gates' action. The gates function based on the amount of weight of these relationships, which are learnt throughout training.

An LSTM-based RNN might be trained supervised on a series of training sequences employing the gradient- descent optimization technique combined with back propagation across time to determine the gradients required for weight changes throughout the process of optimization.

The quantity of each component that should be allowed across is represented by values that the sigmoid layer produces, which are between 0 and 1. "Let nothing across" is indicated by a value of 0, and "let whatever through" is indicated by a value of 1. Three these kinds of gates are present in an LSTM to safeguard and regulate the cell state.

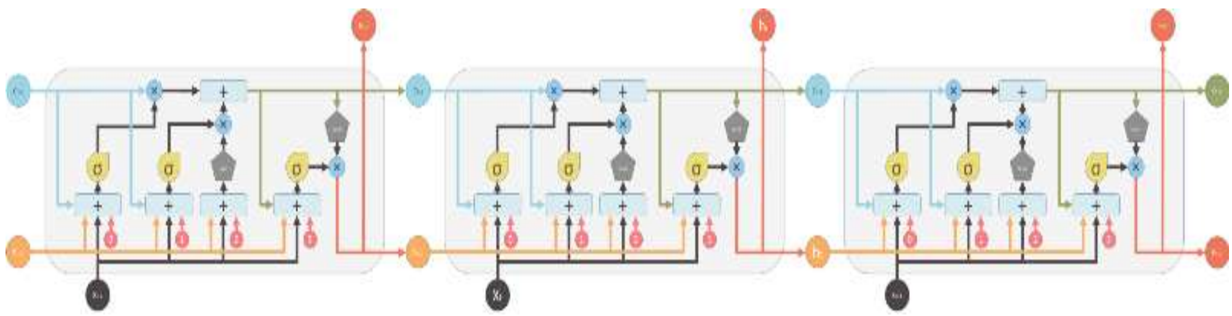


Figure 5 LSTM network with memory blocks

Selecting which data given the cell state to eliminate is the primary step in our LSTM. The sigmoid layer, also referred to as the "forget gate" layer, makes this decision. It examines ht-1 and xt, producing an amount for each cell state that falls between 0 and 1. Ct-1. 0 denotes "totally discard of this," while 1 denotes "totally keep this."

Dataset

Indoor air quality information such as the Air Quality Index (AQI) with respect to hourly and monthly levels about numerous stations crossways numerous towns in India are contained in the Kaggle dataset titled "Air Quality Data in India (2015 – 2020)"1. Using this dataset, the air condition in India is examined from 2015 to 2020, a period of five years1.

This dataset is intended to assist researchers, data scientists, as well as decision-makers in comprehending India's current state of air quality, spotting patterns, and making defensible choices to avoid illnesses brought on by pollutants in the air.

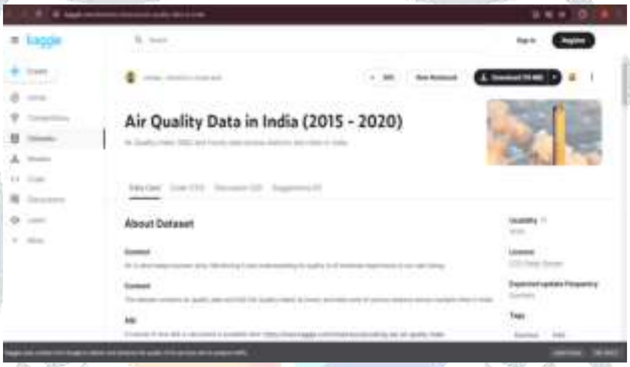


Fig 6 Air Quality Data in India (2015 – 2020)

	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3
count	24172.000000	17764.000000	24463.000000	24459.000000	22993.000000	18314.000000	24405.000000	24245.000000	24043.000000
mean	67.476673	118.454435	17.622421	28.978391	32.289012	23.848366	2.345267	14.362933	34.912885
std	65.075398	89.487976	22.421138	24.627054	30.712855	25.879981	7.075208	17.428655	21.724525
min	0.040000	0.030000	0.030000	0.010000	0.000000	0.010000	0.000000	0.010000	0.010000
25%	29.000000	56.777500	5.660000	11.940000	13.110000	8.960000	0.590000	5.730000	19.250000
50%	48.785000	96.180000	9.910000	22.100000	23.680000	16.310000	0.830000	8.220000	31.250000
75%	80.925000	150.182500	20.038000	38.240000	40.170000	30.360000	1.480000	15.140000	46.080000
max	914.940000	917.080000	390.680000	362.210000	378.240000	352.890000	175.810000	186.080000	257.730000

Table 1 Statistics of Meteorological Data

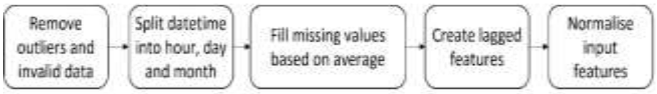


Fig.7. Pre-processing of dataset.

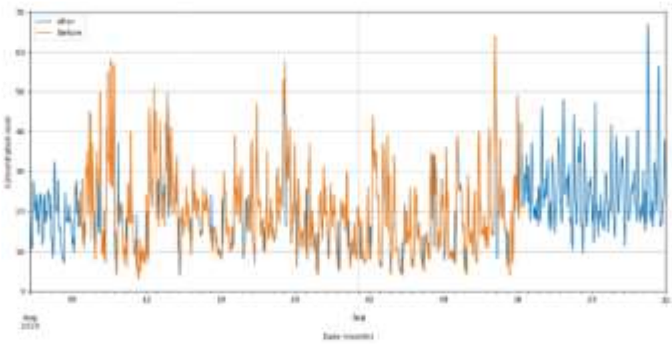


Fig.8. A sample of NO2 data spanning two months, which indicates the insertion of missing values

Results & Analysis

Model Parameters and Tuning: To predict air pollution, the ARIMA model's retraining and related parameters are tested and optimized. In order to forecast each of the five air contaminants in the dataset, we training the model. Yet only NO2 is described as an instance of use for analytical purposes. According to the information as well as the corresponding discrepancies, analysis of ACF as well as partial autocorrelation (PACF) is necessary to determine the ideal values of the model's variables including p, d, and q. We watch the ACF of the real information while ARIMA operates on the stationary data to determine if the dataset is synchronous or non-stationary. The ACF of NO2 is shown in Fig. 9 to be entirely positive and to be steadily declining, indicating that the information is not stationary. We can see how the lag at 1 of 2nd phase distinguishing is negative when we contrast the ACF of the first-order differencing compared to 2nd order differencing. This suggests that the data may become over-differentiated. In contrast, the ACF of initially ordered differencing does not exhibit this behavior and is adequate to achieve data stationarity. Thus, we discovered that the particular sequence of distinction is defined by the value of d, corresponding to 1 (d).

Xnorm = x – xminxmax – xmin (13)

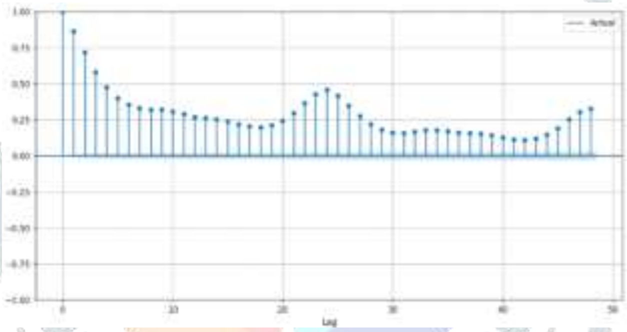


Fig.9. ACF of NO_2 data.

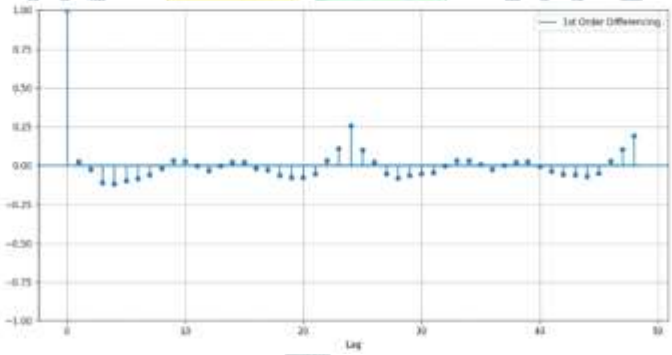


Fig.10. ACF of 1st instruct dissimilarity of NO_2 data

Analogously, by examining ACF cuts off point, a sequence of MA terms depending on significantly differenced data may be discovered. Since the lag at 1 is negative in our instance and also displays the cutoff point, 2nd degree distinguishing indicates an over distinction, resulting in q equal to 1.

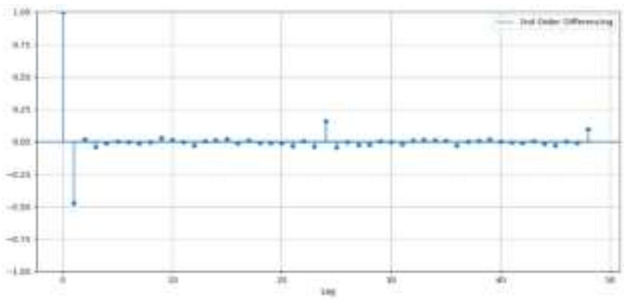


Fig.11. ACF of 2nd order dissimilarity of NO_2 data.

According to the initial cut off point, when PACF at lag 1 is positive, the sequence of AR terms is determined. Figure 6 displays the NO2 PACF data. As illustrated in Fig. 10, the lag at 1 of the first order distinguishing PACF is positive in our instance and indicates the cutoff point, resulting in p equal to 1. Thus, we discovered that the ideal set of variables (p, d, q) over NO2 was (1, 1, 1). Table 2 provides the specifics of the ideal set of settings for the remaining contaminants. To sum up, for any contaminant, the ideal values of d as well as q are determined to be 1. But only p's magnitude is either 0 or 1 for the respective pollutant.

Parameters	Value
No. of layer	1, 1, 5
No. of cells in each layer	5, 5, 30
Dropout layer	0, 0.1, 0.2
Dense layer	1, Relu
Optimiser	Adam
Tuning algorithm	Hyperband

Table 2 synopsis of Deep Learning Model Parameters

Models	Parameters	NO_2	O_3	SO_2	PM2.5	PM10
ARIMA	(p, d, q)	(1,1,1)	(1,1,1)	(0,1,1)	(0,1,1)	(0,1,1)
LSTM	Layers	4	3	2	3	3
	Cells	25,10,15,20	20,10,10	30,5	15,15,25	25,20,15
	Dropout rate	0.0	0.0	0.1	0.1	0.0
GRU	Layers	3	3	2	1	4
	Cells	25,30,15	15,10,10	25,25	20	15,20,15,5
	Dropout rate	0.1	0.0	0.0	0.1	0.1

Table 3 Enhanced Hyperparameters of One-Step Forecasting Equations for Every Pollutant Type

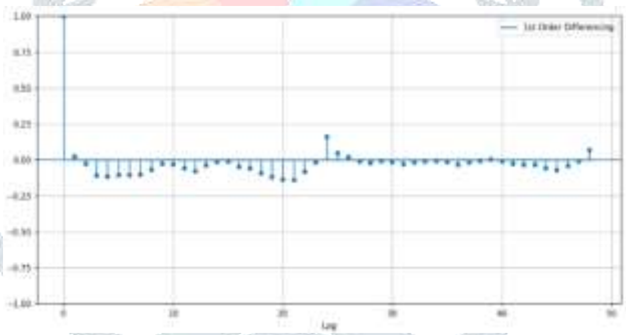


Fig.12. PACF of 1st instruct dissimilarity of NO_2 data

The allocation of the dataset is 70% for training, 20% for justification, and 10% for testing. The indexes in all divisions were maintained more than this in the before batch, preventing shuffling (i.e., improper in time-series). The process of deep learning education and evaluation with essential components is depicted in Fig. 8. For the two types of LSTM and GRU models, Fig. 9 displays the boundaries and stepping size of variables for optimizing the model hyper-parameters with regard to of dropout rate, total amount of layers, and cell size per layer. In deep learning models, the input layer transfers features to an aggregate of five layers, with an absolute minimum of five and a maximum of thirty cells per layer. Our requests are optimizing the amount of layers in DL models in the assortment as of 1 to 5 by a step size of 1. In container of amount of cells, we are allowing for minimum of 5 to maximum of 30 cells by escalating by a step size of 5

Performance Metrics

Metrics called performance indicators are employed to assess a system, manipulate, or activity's efficacy, effectiveness, and success. They are essential for making well-informed decisions to enhance performance and for evaluating effectiveness in relation to goals and objectives. According to the situation, the industry, and the particular objectives, performance measurements can vary greatly, however some popular examples includes:

- Key indicators of performance (KPIs): These particular measures track the advancement of company objectives. Usually, they are measurable and closely related to achievement goals.
- Quality Metrics: These measurements evaluate the caliber of the procedures, goods, or services. Defects and mistake rates, consumer happiness ratings, and standard compliance are a few examples.
- Accounting Metrics: These measures assess a company's profitability by looking at things like revenue, margins of profit, ROI, and expenses per unit.
- Operating Metrics: These measures concentrate on how well and efficiently operational procedures work. Cycle time, productivity, utilization of capacities, and distribution of resources are a few examples.
- Customer metrics: These include customer happiness, rate of retention, net promoter score (NPS), and the lifetime value of a customer (CLV) and are used to quantify multiple facets of the customer interaction.

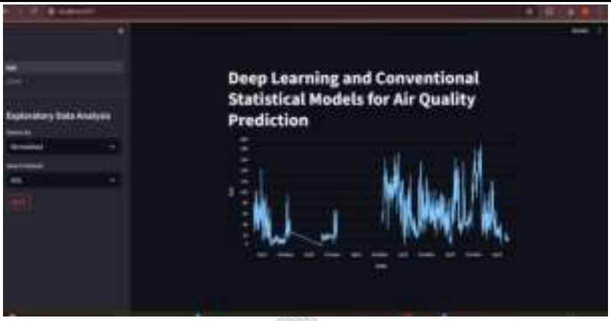


Fig 13. assortment of city and pollutant

We have created Streamlit applications specifically designed to analyze datasets of air quality pollution, with an emphasis on eight different contaminants and certain cities. The above screenshot serves as an example of how the application works. It displays the user interface created to make it easier to explore and analyze information regarding air quality. The program thoroughly examines the air quality and associated pollution levels in every city. The application is designed to enable users to conduct in-depth analysis of pollution patterns and shifts in various places. The screenshot provides an overview of the simple user experience that is intended to enable effective data analysis. Users can learn more about the relationship among pollution concentrations and metropolitan areas by using this program. Complicated data on air quality sets can be accessed and interpreted with ease thanks to the interface. Users may easily access and examine pollution data using the help of interactive graphs and data filters. All things considered, the software provides a strong foundation for comprehending and resolving air quality issues in different cities.

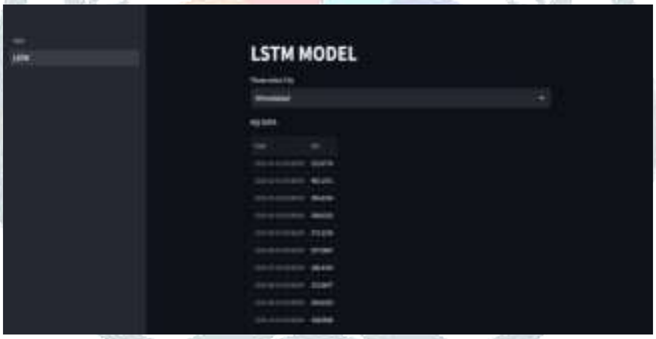


Fig.14. Air quality index of the data with time

The LSTM (Long Short-Term Memory) simulation shown in the following screenshot is intended to forecast particular properties, like SO2, NO2, CO, or AQI (Air Quality Index). A graphical representation of the LSTM model's efficiency may be seen in the image. Different colors are used in this graph to denote different data segments: training data is shown in blue, testing information is shown in green, and future predictions are shown in red. Customers can quickly assess the model's precision at various stages of data processing by using this graphic. The capacity of the LSTM model to produce future forecasts of air quality measurements demonstrates its predictive powers. Interpreting the accuracy of the model during the training, testing, and forecasting stages is made easier by the distinct color separation. The model's capability to identify developments and pattern in the information regarding air quality is easily evaluated by users. The graphical representation makes it easy to appreciate how accurate the model's predictions are over the years. Customers can assess the accuracy of the LSTM algorithm for air quality metric prediction through this graphical illustration. All belongings measured, the screenshot offers perceptive in sequence regarding the predictive capabilities of the LSTM model and its possible use in air pollution level prediction.

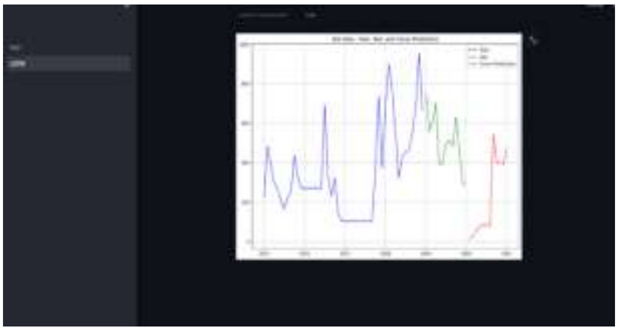


Fig.15. Train, test and prediction of the data

Conclusion

Reducing health hazards and environmental problems is contingent upon effective prediction of air pollution, a worldwide health challenge. The majority of pollutants (for example NO2, O3, SO2, PM2.5, and PM10) are the focus of this work, which uses a variety of methodologies for forecasting employing DL and statistical techniques to predict the impact of pollutants in a single step. R2, MAE, and RMSE are a few examples of evaluation measures that are worn to test the prediction models' effectiveness. Broadly speaking, the ARIMA model finds that the highest RMSE and MAE for NO2 was determined to be 9.354 and 6.065, accordingly, while LSTM attained the lowest RMSE as well as MAE of 0.591 and 0.396, accordingly, across all models for forecasting and pollutants, when assessing SO2 time series data. Out of all the models for forecasting, both DL and models executed comparable in attaining the highest score of approximately 86% while predicting O3. Conversely, the GRU model was shown to have the lowest predicted accuracy for SO2, at about 55%. The general findings from the data showed that, of all the models for projection that were taken into consideration, DL models regularly beat statistical methods in regard to obtaining the least error in regards of RMSE and MAE for every pollutant and superior accurateness of forecasting in regards of R2 for the majority of the pollutants. Although the ARIMA model had the greatest RMSE and MAE error values among all the pollutants, it was only able to forecast two pollutants (i.e., SO2 and PM10) with a higher R2 score. In the coming years, we want to focus on multi-step forecasting and enhance the DL models' effectiveness by utilizing fresh feature engineering techniques and relating optimization of hyper-parameters of the models

Future Scope

Since air pollution is still a major problem in the world, here is a lot of room for development and enhancement in this project. This development could include integrating real-time data streaming through weather stations, orbiting satellites, and sensors, among other sources. The application's integration of real-time data sources allows it to offer instantaneous analysis of the dynamics of air quality, facilitating more prompt and proactive responses. Furthermore, utilizing the latest developments in artificial intelligence (AI) and machine learning methodologies might improve the precision and forecasting powers of the models used in the application. This can entail putting in place more complex computations, like ensemble approaches or deep learning structures, to better identify intricate patterns in information regarding air pollution.

References

1. G. Box, G. Jenkins, G. Reinsel, and G. Ljung, Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2015.
2. M. Stafoggia and T. Bellander, "Short-term effects of air pollutants on daily mortality in the Stockholm County, Sep. 2020.
3. A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ‘‘Air quality forecasting using decision trees algorithms,’ Mar. 2022,
4. B. Wang, W. Kong, H. Guan, and N. N. Xiong, ‘‘Air quality forecasting based on gated recurrent long, short term memory model in Internet of Things,’’2019.
5. B. Paul and S. Louise, Air Quality: Policies Proposals and Concerns—House of Commons Library, Jan. 2022,
6. N. Zaini, L. W. Ean, A. N. Ahmed, M. A. Malek, and M. F. Chow, ‘‘PM2.5 forecasting for an urban area based on deep learning and decomposition method,’ Oct. 2022
7. K. Abutalip, A. Al-Lahham and A. El Saddik, " Digital twin of atmospheric environment: Sensory data fusion for high-resolution PM 2.5 estimation and action policies recommendation ", 2023.
8. J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, Y. Tan, V. J. L. Gan, et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques", J. Cleaner Prod., , Jan. 2020.
9. I. Manisalidis, E. Stavropoulou, A. Stavropoulos and E. Bezirtzoglou, "Environmental and health impacts of air pollution: A review", . 14, 2020.
10. S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, et al., "Comparative analysis of machine learning techniques for predicting air quality in smart cities", 2019.