V Semester Minor Project Report
ON

Air Quality Index Prediction using Deep Learning

Submitted by

Komal Krishna Panigrahi- 191020430
Basharat Khalid Harris- 191020417
Swastik Dashore- 191020455

Under the guidance of
Dr. N. Srinivas Naik

Department of DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

Dr. Shyama Prasad Mukherjee

International Institute of Information Technology, Naya Raipur

(A Joint Initiative of Govt. of Chhattisgarh and NTPC)

Email: iiitnr@iiitnr.ac.in, Tel: (0771) 2474040, Web: www.iiitnr.ac.in

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Dated: December 14, 2021

Komal Krishna Panigrahi (191020430)

Basharat Khalid Harris (191020417)

Swastik Dashore (191020455)

Certificate

This is to certify that the project titled "AIR QUALITY INDEX PREDICTION USING DEEP LEARNING" by *Komal Krishna Panigrahi, Basharat Khalid Harris and Swastik Dashore* has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree/diploma.

Dated: December 14, 2021

Dr. N. Srinivas Naik

Assistant Professor Department of CSE

Dr. SPM IIIT Naya Raipur-493661

December 2021

Approval Sheet

This project report entitled "AIR QUALITY INDEX PREDICTION USING DEEP LEARNING" by *Komal Krishna Panigrahi, Basharat Khalid Harris and Swastik Dashore* is approved for 5th Semester Minor Project-I.

(Signature of Examiner - I)

____

Name of Examiner -I

(Signature of Examiner - II)

____

Name of Examiner -II

(Signature of Chair)

____

Name of Chair

Date: _____________ Place: _____________

Acknowledgments

At the outset, I would like to express my whole hearted and deep sense of gratitude to my guide **Prof. Dr. N. Srinivas Naik** for his guidance, help and encouragement throughout my research work. I greatly admire his attitude towards research, creative thinking, hard work and dedication in work. I am highly grateful to him for patiently checking all my manuscripts and thesis. This thesis would not have been possible without his bounteous efforts. More than a guide, he is my mentor for shaping my personal and professional life, without whom I would not have been where I am today. I owe my profound gratitude to Prof. Dr. N. Srinivas Naik for his support in all respects.

Komal Krishna Panigrahi (191020430)
Basharat Khalid Harris (191020417)
Swastik Dashore (191020455)

Abstract

In recent years, countries all over the world have experienced rapid urbanization. This has resulted in a massive increase of air pollution levels and has created a major concern among the developing countries. Therefore, demand for predicting air quality has become an absolute necessity so that we can take precautionary measures to control the air pollution levels.

Our problem statement here is the analysis of time series meteorological data from various places with different meteorological, temporal and spatial properties to predict and forecast air quality index for the future time with higher accuracy and good performance.

The dataset we have considered consists of major air pollutants as well as the meteorological data. Our plan of approach involves applying Deep Learning models to predict the AQI (Air Quality Index), and in specific, using past AQI and meteorological data to predict the next AQI value. The lesser the difference between the predicted and actual AQI value, the better the model. We will further compare the above mentioned models to deduce the most efficient model out of existing ones.

Table of Contents

List of All Figures

Chapter 1

Introduction

Due to rapid urbanization, many environmental hazards took place in the 20th century, including rise in air pollution levels. Air pollution refers to the introduction of chemicals, particulate matter, or biological matter to the natural environment or atmosphere that cause harm or discomfort to living organisms or damage their life systems. Due to being tiny and light, the air pollutants stay in the atmosphere for a long duration and also easily bypass the filters of the human nose and throat. According to a recent survey, the presence of particulate matter has caused over 4.2 million deaths all around the world.

Particle size is a crucial component in determining the particle deposition location in the human respiratory system. PM2.5, referring to particles with a diameter less than or equal to 2.5 µm, has been an increasing concern, as these particles can be deposited into the alveoli- the lung gas exchange region. The increased association with air pollutants have resulted in increased mortality and morbidity rates . Thus, we have considered PM2.5 as the label for classification here.

The study of AQI can be dated back to the application of statistical models, but it has mostly been restricted to simply utilizing standard classification or regression models, which have not been able to reflect the nature of the problem itself or ignored the correlation between sub-models in different time slots. The development of machine learning and deep learning has helped significantly in refining the model of a specific problem. The model designed in our project helps in predicting the levels of PM2.5 daily with the help of the meteorological data. For achieving this objective, we have applied machine learning algorithms like SVM and Deep learning techniques like ANN and LSTM. The model built then compares and analyzes the results using the evaluation metrics and mean squared error. The model also helps us in determining the correlation between various meteorological parameters and PM2.5.

Chapter 2

Literature Survey

AQI prediction in smart cities has been done through various ML and DL models and still remains a challenging problem. Several deep learning methods comprise of long short-term memory (LSTM) network models which also include some recent versions such as bidirectional-LSTM and encoder-decoder LSTM models use a multivariate time series approach that attempts to predict air quality for 10 prediction horizons covering a total of 80 hours and provide a long-term (one month ahead) forecast.[15]. Some others use GRU as a replacement of LSTM as it is expected to be less complex than it. GRU usually performs better than LSTM except in some cases of outliers. Noisy data makes the model worse. Various big-data and machine learning-based techniques are used for air quality forecasting. [14]Attempts have been made for weather forecasting using spatial-temporal relations. A combination of multiple neural networks, including an artificial neural network, a convolutional neural network, and a long-short-term memory to extract spatial-temporal relations. The model with all components (Adaptive+LSTM+Convolution Neural Networks) outperformed all other models, and the CNN only model outperformed all other combinations for 2-6 hour predictions[13]. But the approach couldn't make the model robust. Similar other models such as CNN were used first to efficiently extract features from the data, and then LSTM was adopted to predict the air quality with the extracted feature data. The CNN-LSTM model was compared to MLP, CNN, RNN, LSTM, and CNN-RNN using the same training and test set data in the same operating environment to demonstrate its effectiveness[11]. The only problem with this model is that it is unable to produce a precise prediction for the Long term. Another approach utilised Long Short Term Memory (LSTM) and Gated Recurrent Unit(GRU) networks to predict future values of air quality. GRU network was employed as the model's first hidden layer and LSTM as the model's second hidden layer, followed by two dense layers[10]. It still performs better than previous models but noisy data makes it less effective. Hybrid ensemble models such as CERL exploit the merits of both forward neural networks and recurrent neural networks that are designed for handling time serial data to predict air quality hourly[16]. There are many studies proposing solutions for the prediction problem of AQI in the literature. These methods are physical deterministic models, statistical, empirical and neural network approaches that can be categorized in deterministic and statistical methods[17].

Overview of LSTM and GRU

RNN is the greatest solution for language models since it can maintain long-haul conditions. However, slopes of RNN may vanish as delays increase while the RNN unfurls into incredibly deep feed neural architecture. Because the evaporating slopes problem is so serious, LSTM and GRU were given neglect door units to deal with it. These units allowed memory cells to close when data needed to be ignored. That is how it determines appropriate time delays in this manner. The construction of LSTM and GRU is briefly examined in the two subsections that follow.

LSTM

For building the language-based models, "LSTM" was presented at first back in
1997. Since this long momentary memory (LSTM) has magnificent ability to recollect
Long term faithfulness. Four kinds of gates are accessible in old-style LSTM. They are:

1. Input Gate
2. Input Modulation Gate
3. Forget Gate
4. Output Gate.

The structure of a typical LSTM cell is shown in Fig. 1 [18].

Input gates responsibility is to deal with every one of the new information which comes from the rest of the world. The result of the information adjustment door goes to the memory cell. In the accompanying emphasis, the neglected door chooses the data which to keep and which not.By doing this, it picks the ideal deferrals for the info information succession. The outcomes which are determined go as contributions to the resulting door. The resulting door produces the result of the long momentary memory cell. Generally, a softmax layer is stacked on top of the result layer of the LSTM in language models. Be that as it may, in our model, a thick layer is stacked on top of the result layer of the LSTM cell.In this procedure,X = (x1,x2,...,xn) is the input time series,H= (h1,h2,...,hn) is hidden state of memory cells, Y=(y1,y2,...,yn) is output time series. The hidden state of the memory cell is calculated using this formula:

$$ht = H(W_{hx}\, x_t + W_{hh} h_{t-1} + b_h)$$

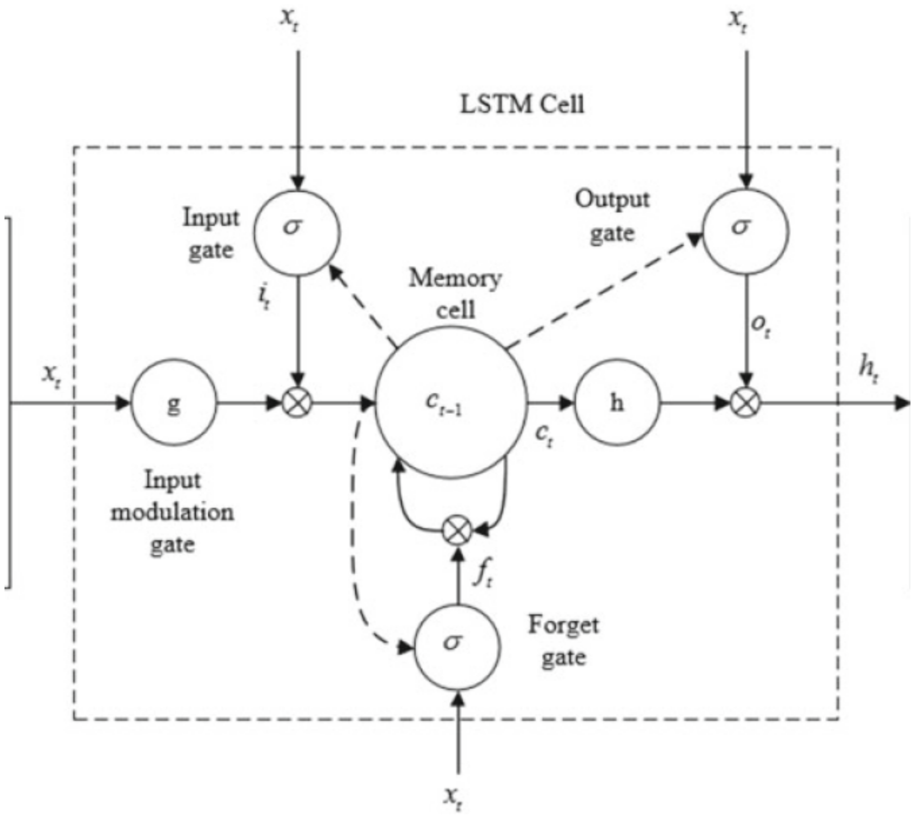


Fig 2.1 LSTM cell

Here, weight matrices are denoted as "W", and bias vectors are denoted as "b". And by using the following formulas, the hidden state of the memory cells is computed

$$p_t = W_{hy}\, y_{t-1} + b_y$$

$$i_t = \sigma(W_{ix}\, x_t + W_{hh} h_{t-1} + W_{ic} c_{t-1} + bi\ )$$

$$f_t = \sigma(W_{fx}\, x_t + W_{hh} h_{t-1} + W_{fc} c_{t-1} + b_f\ )$$

$$c_t = f_t * c_{t-1} + i_t * g(W_{cx}\, x_t + W_{hh} h_{t-1} + W_{cc} c_{t-1} + b_c)$$

$$o_t = \sigma(W_{ox}\, x_t + W_{hh} h_{t-1} + W_{oc} c_{t-1} + b_o)$$

$$h_t = o_t * h(c_t\ )$$

Backpropagation through time (BPTT) is done with the Adam optimizer to avoid local minima and decrease training mistakes [19]. Overfitting is a problem with neural networks. To address the problem of overfitting, a large number of regularisation approaches are presented.

GRU

In 2014, Cho et al. introduced a much simpler version of RNN which is called GRU [20]. The shape of a typical GRU cell is illustrated in the following Fig. 2 [11].

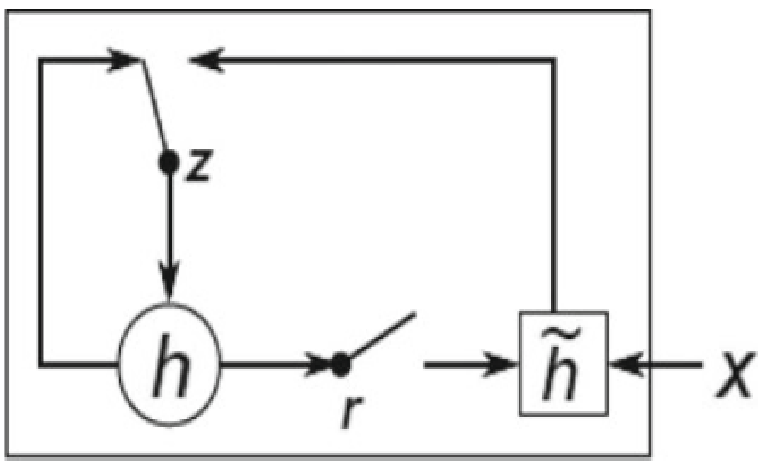


Fig 2.2 GRU cell

In a general GRU cell, there are two gates. The reset gate (r) is the first, and the update gate (u) is the second (z). The result of the hidden state is computed at time t using the value of the hidden state result at time t-1 as well as the input time series value at timestamp t, which is fundamentally similar to the LSTM cell. The hidden state yield at time t is calculated using the hidden state at time t-1 and the input time series esteem at time t, which is similar to that of an LSTM cell. Furthermore, the reset entryway of the GRU cell works similarly to the forget gate of a long short term memory cell.

SARIMA

SARIMA is among one of the most popular and classical models for Time series forecasting. It consists of three components(AutoRegression, moving average, Integrated). A pure **Auto-Regressive (AR only)** model is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$\mathbf{Y_t = \alpha + \beta_1 * Y_{t-1} + \beta_2 * Y_{t-2} + \ldots\ldots + \beta_p * Y_{t-p} + \varepsilon_1}$$

where, Y{t-1} is the lag1 of the series, beta1 is the coefficient of lag1 that the model estimates and alpha is the intercept term, also estimated by the model.Likewise a pure Moving Average (MA only) model is one where Yt depends only on the lagged forecast errors.

$$. \mathbf{Y_t = \alpha + \varepsilon_t + \phi_1 * \varepsilon_{t-1} + \phi_2 * \varepsilon_{t-2} + \ldots\ldots\ldots + \phi_q * \varepsilon_{t-q}}$$

An ARIMA model is one where the time series was differentiated at least once to make it stationary and you combine the AR and the MA terms. So the equation becomes:

$$\mathbf{Y_t = \alpha + \beta_1 * Y_{t-1} + \beta_2 * Y_{t-2} + \ldots\ldots + \beta_p * Y_{t-p} \quad + \quad \varepsilon_t + \phi_1 * \varepsilon_{t-1} + \phi_2 * \varepsilon_{t-2} + \ldots\ldots\ldots + \phi_q * \varepsilon_{t-q}}$$

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (up to q lags)

Chapter 3

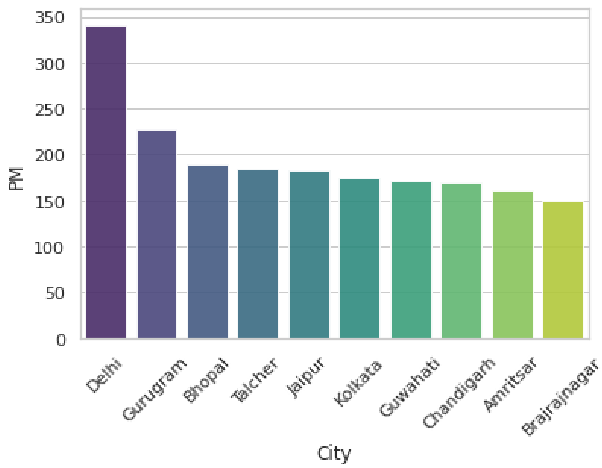
Insight into our Dataset:

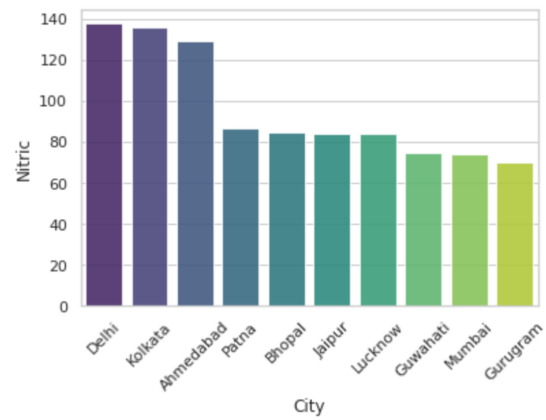
The dataset is extracted from the **Central Pollution Control Board of India(CPCB)**. It consists of AQI indices of various cities of India on a daily basis from 2015 - 2020. It also contains concentration of various pollutants which is taken into consideration while calculating AQI.A snapshot of the data is given below.

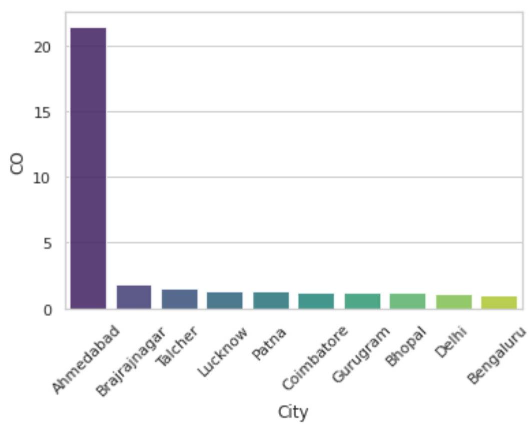
	City	year	month	PM	Nitric	CO	NH3	O3	SO2	BTX	AQI
0	Ahmedabad	2015	1	10.668710	88.680000	22.352258	0.0	46.350645	43.602903	6.971613	33.903226
1	Ahmedabad	2015	2	103.662143	92.985714	19.482143	0.0	43.437857	56.423214	35.357143	464.857143
2	Ahmedabad	2015	3	106.905806	80.510000	13.585484	0.0	44.276774	56.975161	41.357419	378.064516
3	Ahmedabad	2015	4	101.682000	54.992667	7.306333	0.0	31.376000	51.233333	14.496333	257.200000
4	Ahmedabad	2015	5	74.919355	50.607419	8.529677	0.0	31.624194	35.977419	19.677419	254.967742

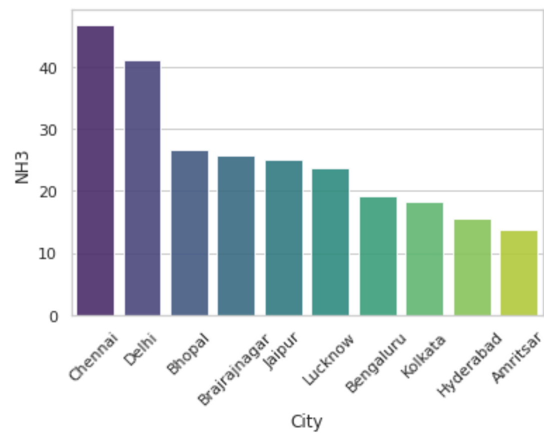
Fig 3.1 A snap of dataset

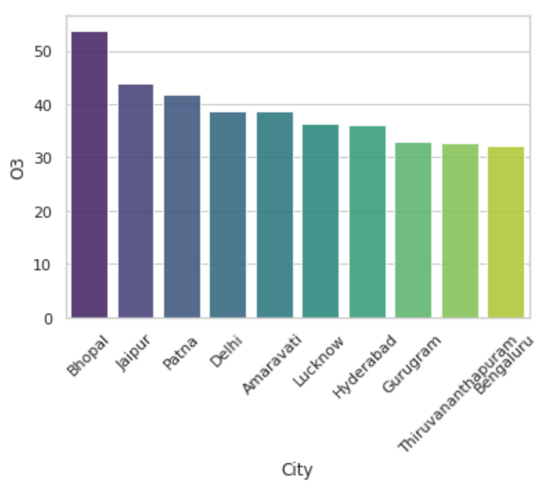
The concentration of various pollutants in different cities is depicted by following histograms:

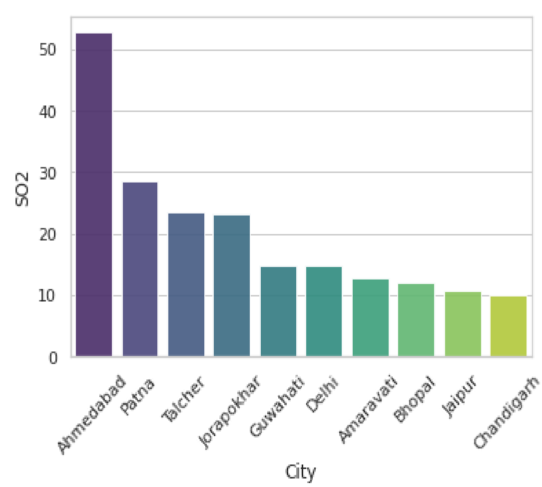

PM
0
50
100
150
200
250
300
350
Delhi
Gurugram
Bhopal
Talcher
Jaipur
Kolkata
Guwahati
Chandigarh
Amritsar
Brajrajnagar
City


Nitric
0
20
40
60
80
100
120
140
Delhi
Kolkata
Ahmedabad
Patna
Bhopal
Jaipur
Lucknow
Guwahati
Mumbai
Gurugram
City


CO
0
5
10
15
20
Ahmedabad
Brajrajnagar
Talcher
Lucknow
Patna
Coimbatore
Gurugram
Bhopal
Delhi
Bengaluru
City


NH3
0
10
20
30
40
Chennai
Delhi
Bhopal
Brajrajnagar
Jaipur
Lucknow
Bengaluru
Kolkata
Hyderabad
Amritsar
City


O3
0
10
20
30
40
50
Bhopal
Jaipur
Patna
Delhi
Amaravati
Lucknow
Hyderabad
Gurugram
Thiruvananthapuram
Bengaluru
City


SO2
0
10
20
30
40
50
Ahmedabad
Patna
Talcher
Jorapokhar
Guwahati
Delhi
Amaravati
Bhopal
Jaipur
Chandigarh
City

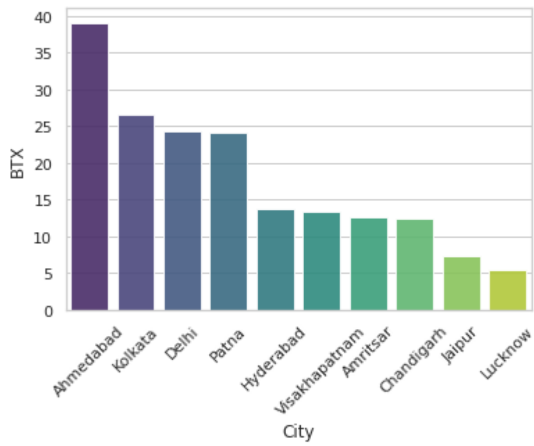


Fig 3.2 depicts the concentration of various pollutants in different cities.

Below is the correlation matrix for finding the relationship between each pollutant and AQI index.

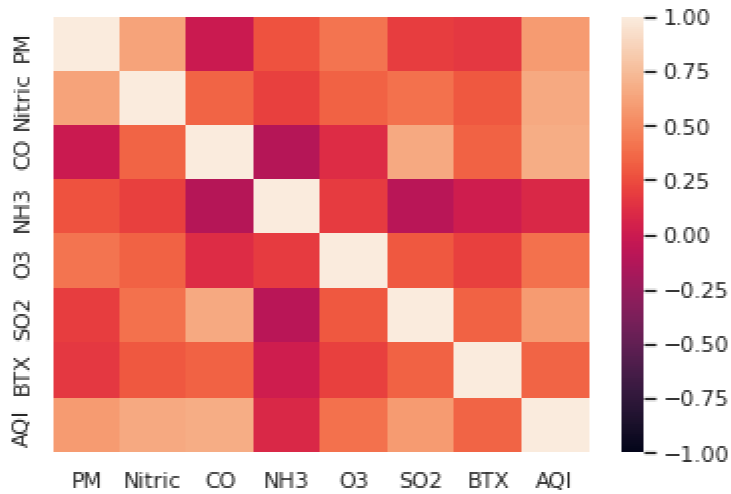


Fig 3.3 displays the correlation between different pollutants and AQI

We see that BTX has the lowest correlation with AQI- which is perfectly in sync with the AQI calculation formula. The air quality index is composed of 8 pollutants ((PM10, PM2.5, NO2, SO2, CO, O3, NH3, and Pb), but does not directly account for BTX.

A general AQI trend over the months from year 2017-19:

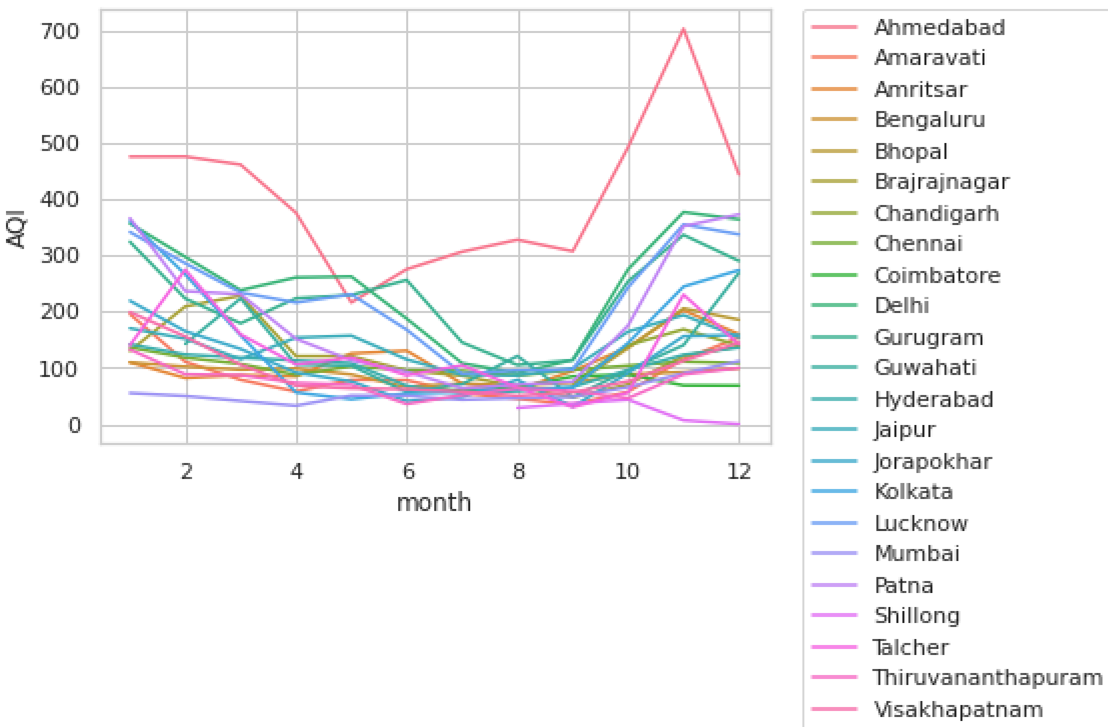


Fig 3.4 depicts the variation of AQI indices over months of 2019 for various cities

Emergence of a clear pattern is observed. AQI decreases in the summer months, which in turn means that air quality improves over these months.

Also, a significant improvement in AQI can be seen over a few months of 2020.A possible reason could be the effect of lockdown.The general trend shows that the AQI indeed decreased for the lockdown months, signifying a major improvement in Air quality with reduced pollution levels.

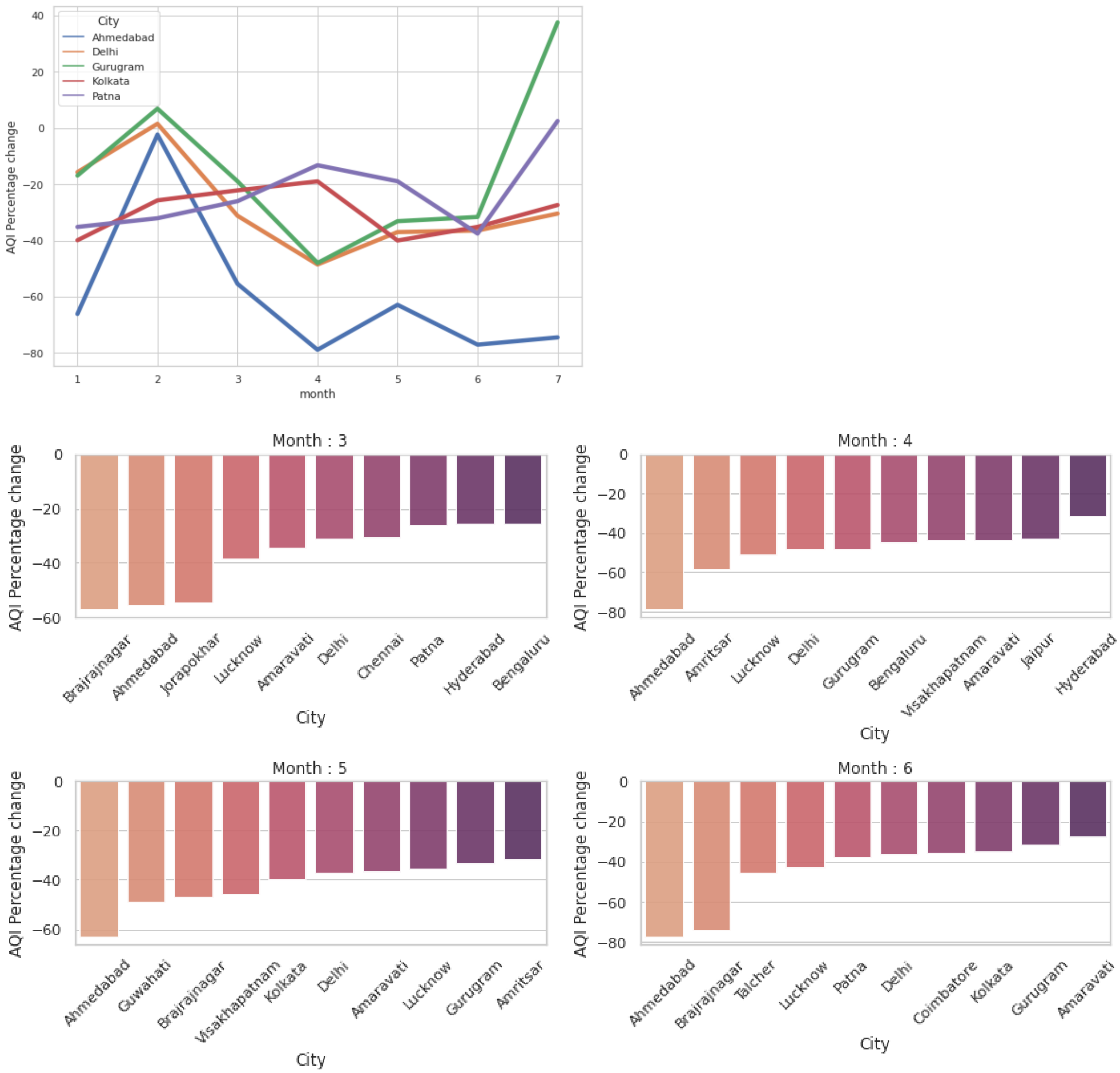


Fig 3.5 depicts the percentage change in AQI for year 2019 over few months for different cities

We can see that there has been a significant improvement in the air quality for these cities over the four months.

A complete plot of mean AQI indices over all cities from 2015-2020 can be shown below:

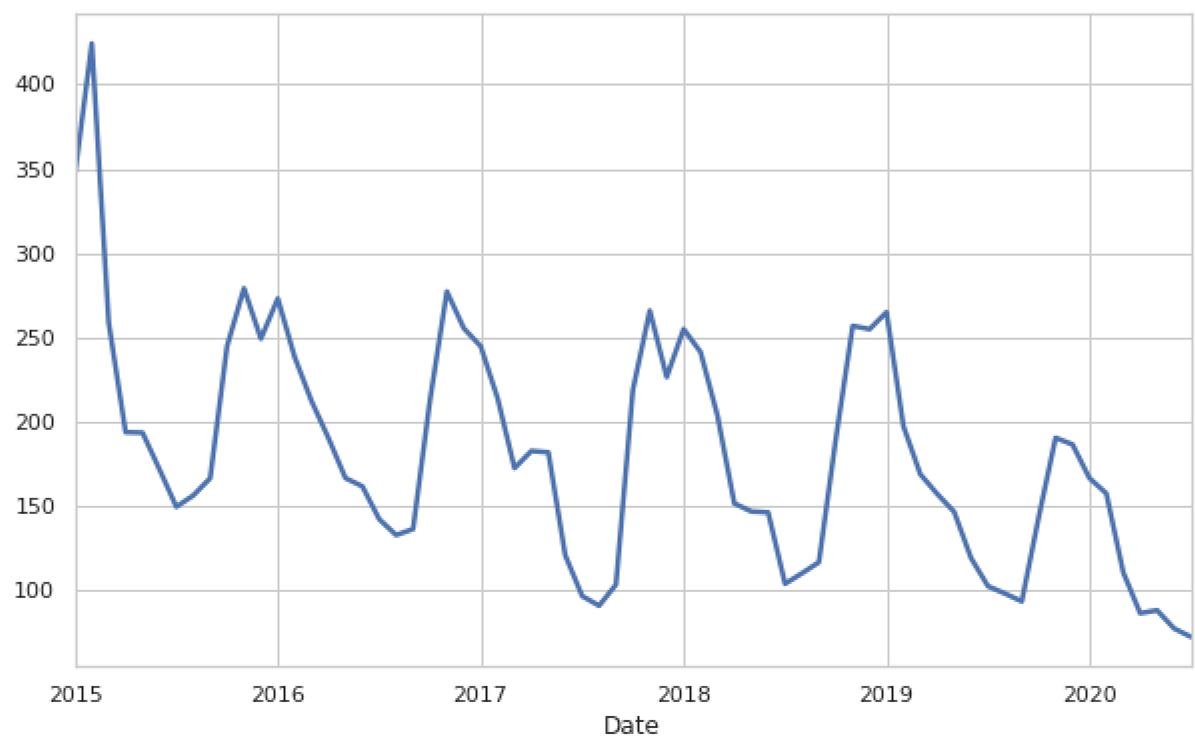


Fig 3.6 plot of mean AQI indices for various cities over the year 2015-2020

From observation of the above fig, it can be seen that the time series comprised various characteristics such as repeated pattern or cycle. The various component of the time series can be decomposed into its components below:

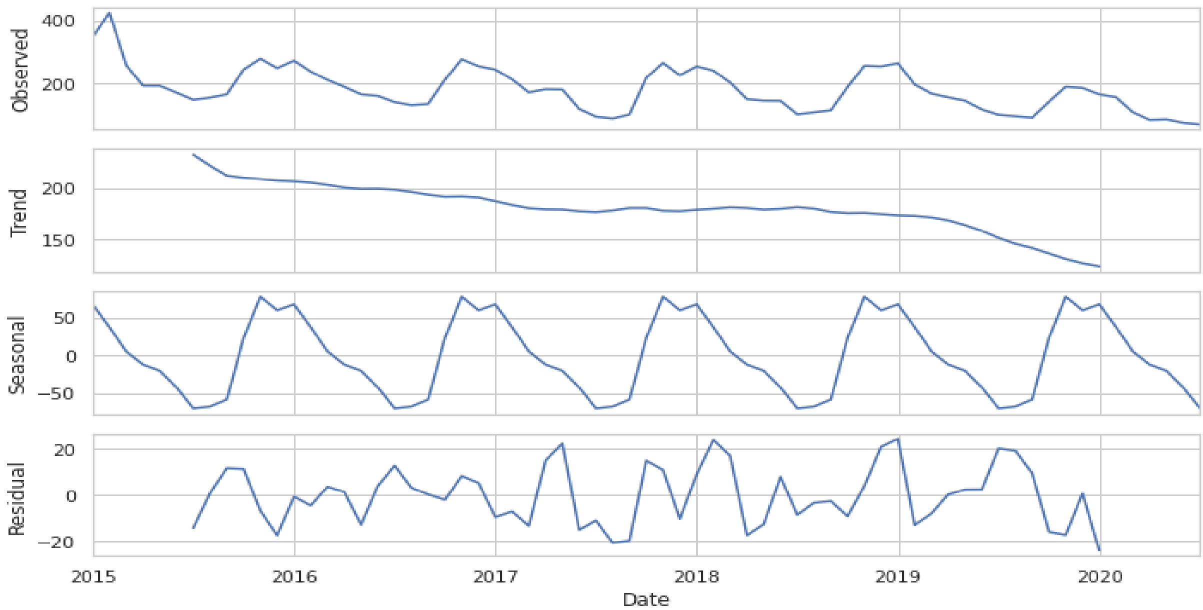


Fig 3.7 decomposition of original time data into its components of seasonality.trend and residuals

Above components suggest that there is a seasonality in the data as indicated by the seasonal component.Trend signifies a non-constant mean over the years from 2015-2020 which is a clear indication that time series is not stationary.

Also, it's important to plot the correlation and autocorrelation between the time lagged versions of the AQI index.

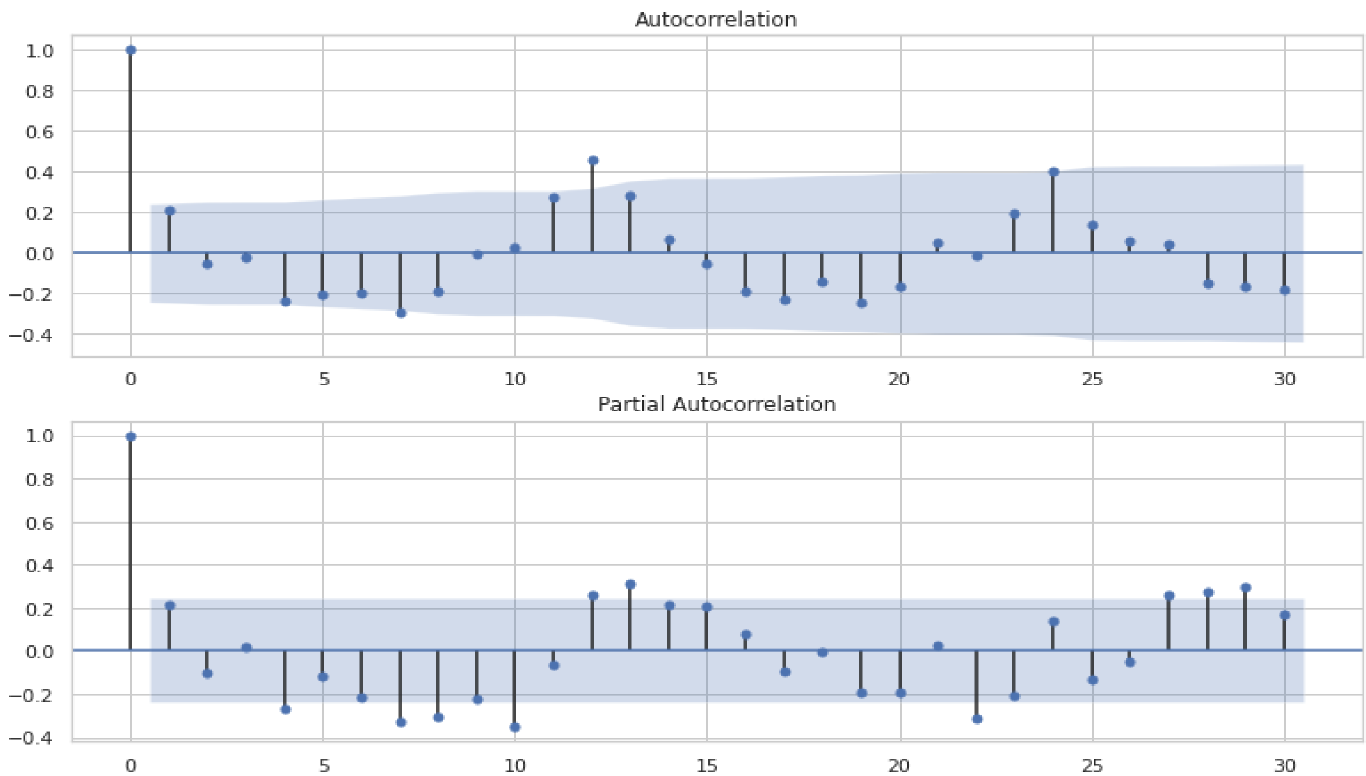


Fig 3.8 Plot of Autocorrelation and Partial Autocorrelation for different time lagged versions of AQI

Chapter 4

Software Requirement Specification:

- **Tensorflow:** It is a free, symbolic math and open source software library for machine learning which is based on differential programming and dataflow. It was developed by Google Brain, for the variety of tasks among which the important ones are training and inference of deep neural networks.We have used this framework, since it can work on our large scale data and is great in numerical computation.

- **Keras:** It is a deep learning framework built on top of Tensorflow, written in python to do fast experimentations and quickly obtain the results. We used it to take advantage of the several implementations in it, such as Convolutional layers, batch normalization,pooling as well as the Adam optimization function.

- **Numpy:** It is a library used for working with multidimensional arrays and matrices, and supports many high level mathematical functions. Numpy is a short form of Numerical Python. We have used numpy to increase the dimension of images using expand_dims() function.

- **Pandas:** It is a python library used for reading, analysis and manipulation of data. We can work with various file formats using pandas such as .csv,.json etc. It was used for reading the json files while we were downloading the data.

- **Matplotlib:** It is a python library used for visualization of data.

Chapter 5

Proposed Solution :

The proposed solution works in some stages as depicted in the flow chart. First Data collection and preprocessing is done. Data is collected from the CPCB portal for the national AQI index for various cities from Jan 2015 to July 2020.

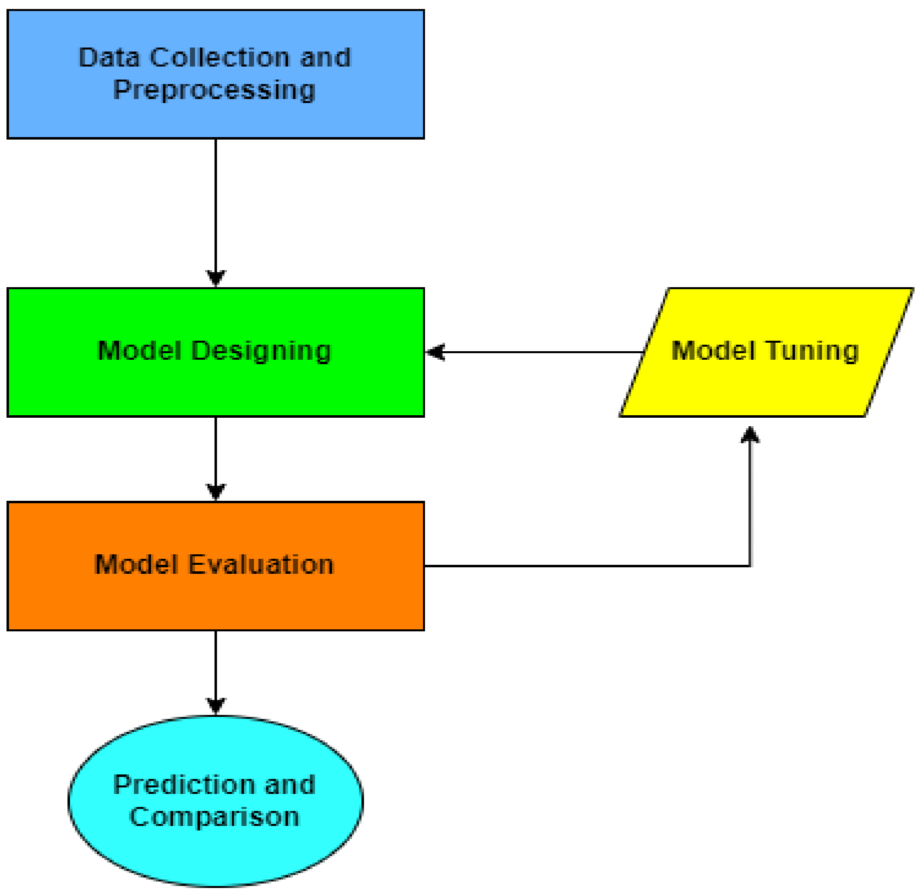


Fig 5.1 Workflow diagram

Data Collection and Preprocessing

Data is taken from the CPCB portal for various cities for the duration of Jan 2015 to July 2020 on a daily basis. The dataset contains levels of various pollutants in the air and the AQI values along with the AQI bucket. For our task, we have pivoted the table for cities AQI and dates. We determined the Average AQI for all cities for each particular day. It is termed India_AQI. Mean imputation is performed on the data using a central average of window size 20. The data is Normalized using minMaxScaler. The Task is to perform univariate time series prediction and forecasting.

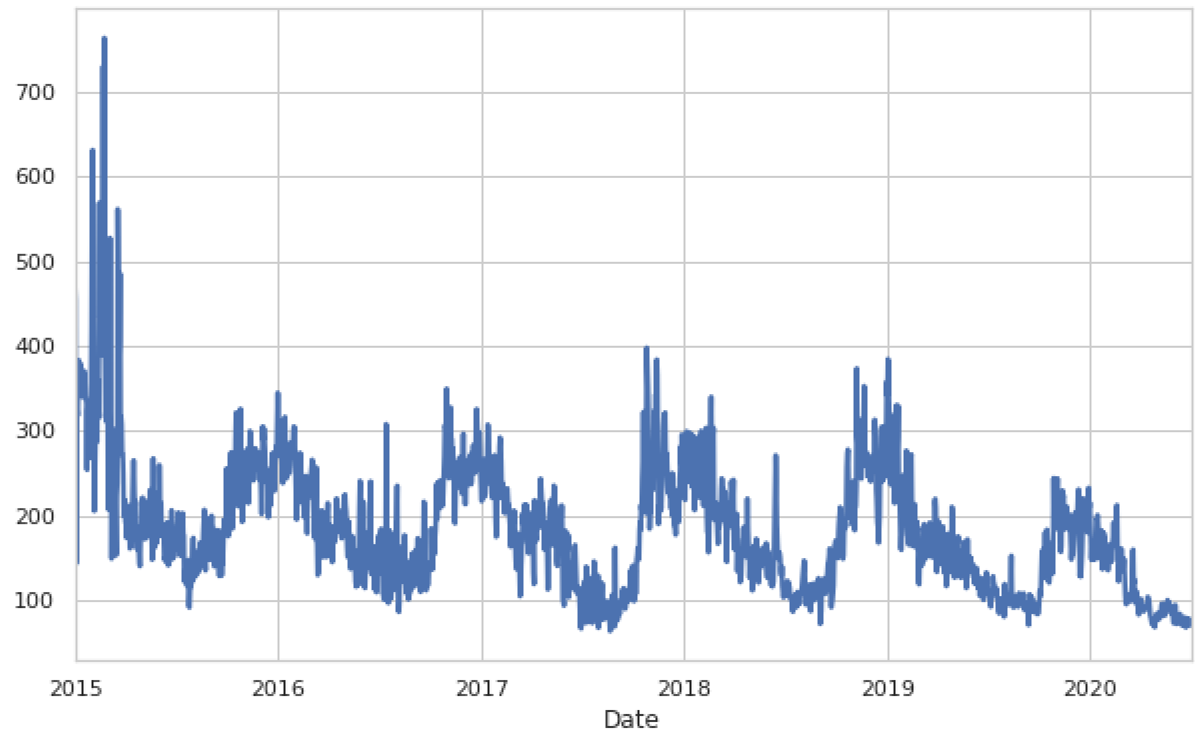
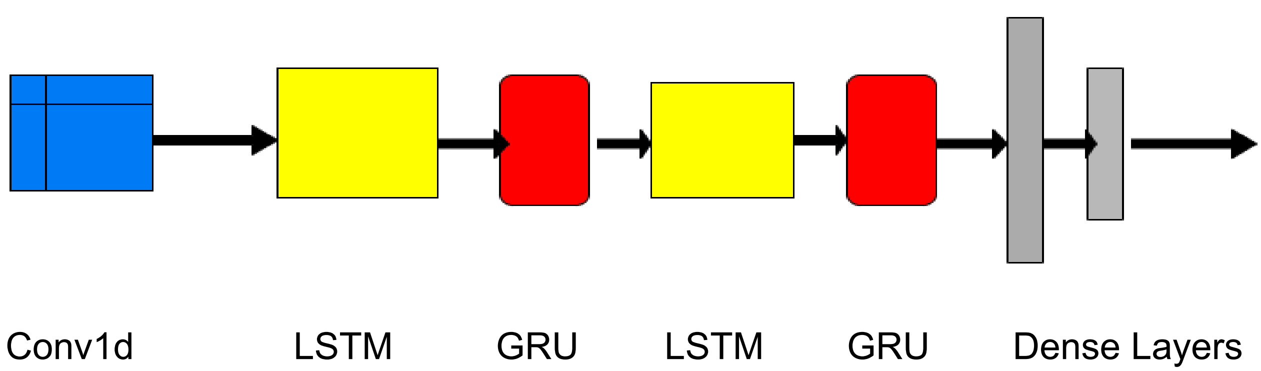


Fig 5.2 A plot of the complete dataset.

Model Designing

The proposed model consists of 4 types of layers: conv1d layer, alternating LSTM and GRU units followed by Dense layers and Lambda layer. The first layer is conv1d with 60 filters and kernels of size 5. It is followed by LSTM units with 64 cells. Then 64 units of LSTM cells are used. Again a similar LSTM and GRU part is repeated. It is then followed by a Dense layer of 30 units. Then it is followed by another dense layer of 10 units and then a single Dense unit.



The Conv1d layer helps in extracting features and patterns from the Time series. Then the combination of LSTM and GRU helps in making predictions for the future data. Optimizer used is stochastic Gradient Descent with momentum. Loss function used is the Huber loss function. The huber loss function is a speciality of handling outliers. It penalizes large outliers. The model predicted output is compared against the actual output. This process is repeated for 200 epochs, while the weighted values are updated to reduce the value of the cost function (mean absolute error).

Model Evaluation Criteria

Assessing the performance of the model is significant since it is utilized to gauge how the model acts. For this purpose, we have utilized mean absolute error (MAE), mean squared blunder (MSE), and root mean squared mistake (RMSE) to approve the execution of our proposed model. Equations below show the conditions for these Errors.

$$\text{MSE} = \frac{1}{n}\sum_{j=1}^{n}\left(y_i - \hat{Y}_i\right)^2$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\sum_{1=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n}}$$

MAE computes the overall degree of forecasting errors, while MSE and RMSE put more weight on huge mistakes as they square the errors prior to averaging. The more modest The value of these measurements are, the better the model's expectation.

Chapter 6

Experiments and Result Analysis:

6.1 Result and Analysis

We performed a prediction and forecasting using the SARIMA model by taking years on the x-axis and the mean AQI indices of various cities on the y-axis. For testing, we trained the model on a dataset from the year 2015-2019 and tested on the year 2020 data..The test results gives an RMSE of 22.755 and a mean absolute error of 18.044.

For deep learning models, we created a windowed dataset with a window size of 40. We created training and testing sets with a batch size of 100. The First 1500 observations were taken in the training set and remaining in the test set. Then we fit our data on models with only LSTM in which we get MAE as 61.31. Also, fit data on a model with only GRU we get MAE as 53.82. Combined model with LSTM and GRU gave an MAE of 58.03.

In contrast, our proposed model outperformed all of them and gave a MAE of 14.666. Our model is capable of performing well in very noisy data. A plot of its predictions is given below.

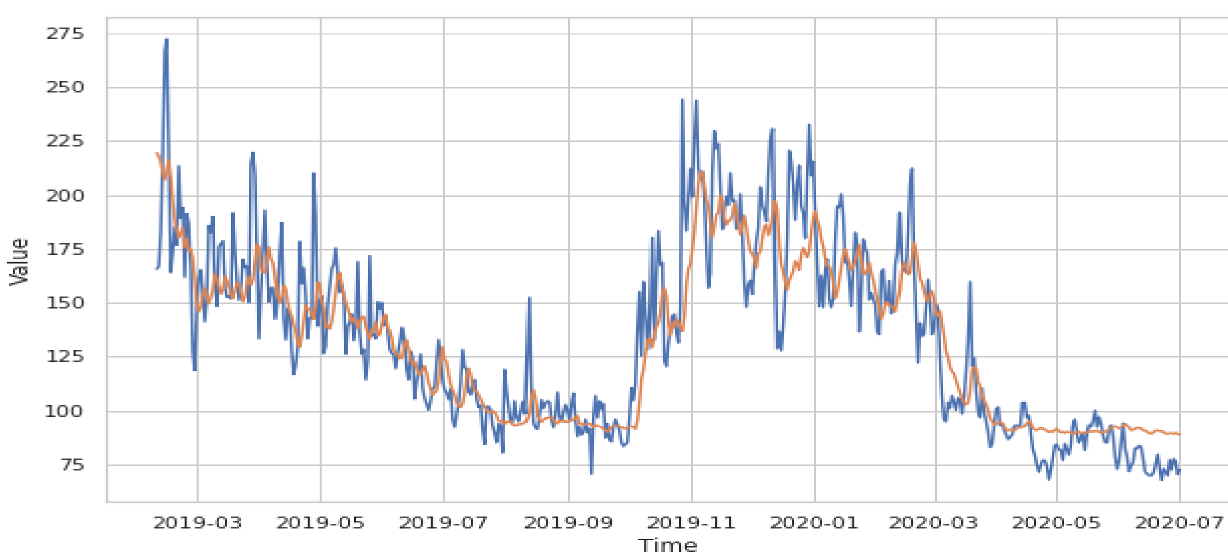


Fig 6.1 Plot of test results of prediction on the year 2020 with Proposed model

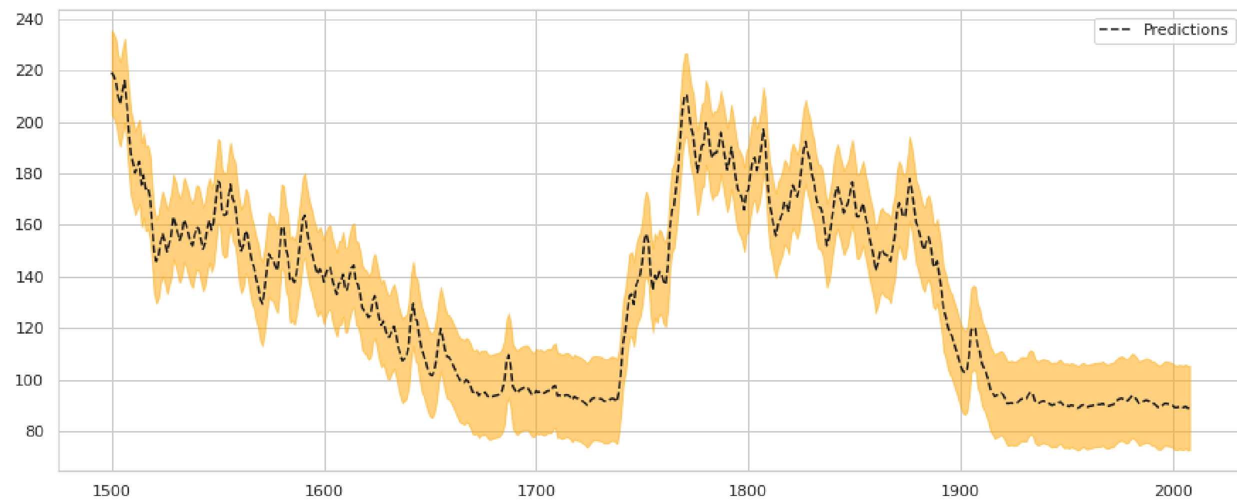


Fig 6.2 Plot of forecasting with confidence intervals produced by proposed model.

6.2 Comparison with the existing solutions

As mentioned in the Literature Review Chapter, most of the past solutions which used heavy deep learning architectures used a lot of memory space and hence were unable to perform real time analysis. Also, those solutions were not able to perform well with the long term predictions.

In our prototype, we gave a robust model capable of performing well on noisy data. The model outperformed previous models as the data taken is highly noisy. The below table shows a comparison of some previous models and our proposed model.

	MAE	MSE	RMSE
Only LSTM	61.31979	4566.3457	67.5747
Only GRU	53.826256	4137.462	64.3231
LSTM + GRU	58.038548	4560.7285	67.5331
Proposed Model	14.666262	450.5907	21.2271

Table 6.1 Comparison of performance of various models

Chapter 7

Conclusion and Future Scope:

In this work, we proposed a hybrid deep learning model for AQI forecasting on the Central Pollution Control Board of India (CPCB) dataset, which consists of AQI indices of various cities of India on a daily basis from 2015 - 2020. It also contains the concentration of various pollutants which is taken into consideration while calculating AQI. We used various memory units such as LSTM, GRU and Conv1d for this time series. We compared our proposed model with existing models and concluded that our model is well suited in a noisy environment.

Since our test data is long and noisy, the proposed model performed well. The model is quite inefficient in terms of memory and may sometimes lead to overfit in extremely regular time series data. Thus the reduction of complexity still remains an open problem in this domain.

Chapter 8

References:

1. [1] X. L. L. P. Y. H. J. S. T. Chi, "Deep learning architecture for air quality predictions," Environ Sci Pollut Res, vol. 23, pp. 22408–22417, 2016.
2. [2] R. S. P. Awkash Kumar, Anil Kumar Dikshit, Rakesh Kumar, "Application of AERMOD for short-term air quality prediction with forecasted meteorology using WRF model," Clean Techn Environ Policy, vol. 19, pp. 1955–1965, 2017.
3. [3] Z. Yang and J. Wang, "A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction," Environmental Research, vol. 158, pp. 105-117, 2017.
4. [4] J. Wang, X. Zhang, Z. Guo, and H. Lu, "Developing an early-warning system for air quality prediction and assessment of cities in China," Expert Systems with Applications, vol. 84, pp. 102-116, 2017.
5. Athira V, Geetha P, Vinayakumar R, Soman K P,
   DeepAirNet: Applying Recurrent Networks for Air Quality Prediction,
   Procedia Computer Science, Volume 132, 2018, Pages 1394-1403,
   ISSN 1877-0509
   https://www.sciencedirect.com/science/article/pii/S1877050918308007
6. Zhao, Z., Qin, J., He, Z. et al. Combining forward with recurrent neural networks for hourly air quality prediction in Northwest of China. Environ Sci Pollut Res 27, 28931–28948 (2020). https://doi.org/10.1007/s11356-020-08948-1

   https://link.springer.com/article/10.1007%2Fs11356-020-08948-1
7. Wang, J., Li, J., Wang, X. et al. Air quality prediction using CT-LSTM. Neural Comput & Applic 33, 4779–4792 (2021). https://doi.org/10.1007/s00521-020-05535-w

   https://link.springer.com/article/10.1007/s00521-020-05535-w
8. https://www.hindawi.com/journals/complexity/2020/8049504/#abstract
9. "Author index," Proceedings HPCA Seventh International Symposium on High-Performance Computer Architecture, 2001, pp. 317-318, doi: 10.1109/HPCA.2001.903273.
10. A Novel Deep Learning Approach to Predict Air Quality Index By- Emam Hossain, Mohd Arafath Uddin Shariff, Mohammad Shahadat Hossain and Karl Andersson

11. A CNN-LSTM-Based Model to Forecast Air Quality Wenjie Lu, Jiazheng Li, Yifan Li, Aijun Sun and Jingyang Wang.
12. A Deep Learning Model for Air Quality Prediction in Smart Cities - İbrahim KÖK ,Mehmet Ulvi ŞİMŞEK, Suat ÖZDEMİR
13. Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations - Ping-Wei Soh, Jia-Wei Chang AND Jen-Wei Huang
14. Air Quality Prediction: Big Data and Machine Learning Approaches - Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu and Gang Xie.
15. Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown -Animesh Tiwari, Rishabh Gupta, Rohitash Chandra
16. Combining forward with recurrent neural networks for hourly air quality prediction in Northwest of China - Zhao, Z., Qin, J., He, Z
17. Studies proposing solutions for the prediction problem of AQI in the literature
18. Fu, R., Zhang, Z., Li, L.: Using lSTM and GRU neural network methods for traffic flow prediction.In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation(YAC), pp. 324–328. IEEE (2016)
19. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprintarXiv:1412.6980 (2014)
20. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio,Y.: Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078 (2014)