A comprehensive study to predict the short-term air quality with the help of multiple deep learning models in Guangzhou city: China

MSc Research Project
MSc Data Analytics

Virender Singh
Student ID: X17150671

School of Computing
National College of Ireland
Supervisor: Dr. Anu Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Virender Singh……..

Student ID: X17150671……………………………………………………………………………

Programme: MSc Data Analytics…………………………… **Year:** 2018-2019

Module: Research in Computing……………………………………………………

Supervisor: Dr. Anu Sahni……………………………………………………………

Submission Due Date: 12th Aug,2019………………………………………………………………

Project Title:

A comprehensive study to predict the short-term air quality with the help of multiple deep learning models in Guangzhou city: China

Word Count: 6198 **Page Count: 26**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Virender Singh……………………………………………………………………

Date: 12th Aug,2019……………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	◻
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	◻
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	◻

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A comprehensive study to predict the short-term air quality with the help of multiple deep learning models in Guangzhou city: China

Virender Singh

X17150671

Abstract

Air pollution is recognised as a worldwide risk to humans. Large cities in China particularly have experienced a marked degradation in air quality so residents are at particular risk. Prediction of the immediate future air quality is important as accurate warnings could assist residents and industries take measures to reduce exposure to and production of pollution when meteorological conditions are expected to increase the impact of pollutants. The research aims to establish whether patterns in the historic data can assist identify sets of occurrences that might combine to negatively impact air quality. The focus of this research is on levels of the principal pollutant, particulate matters 2.5 (pm2.5). These particles are particularly dangerous because of their tiny size, with a diameter of 2.5 microns or less these are the most significant cause of air quality related deaths. This research tested the following prediction methods; Convolutional Neural Networks (CNN), Gated Current Units (GRU), Multilayer perceptron (MLP) and Long-Short Term Memory (LSTM). A performance comparison completed with the historically executed Machine Learning (ML) models using the CRISP-DM methodology demonstrated that the methodologies can successfully predict the concentration level of pm2.5 and the correlation with meteorological factors that leads to air quality degradation.

Keywords: air pollution, LSTM, GRU, CNN, MLP, pm2.5

1. Introduction:

Urban air pollution is an influencer on health. This research focuses firstly on executing prediction models for short term predictions of pollution levels to help avoid further harm to the already choking city: Guangzhou (Yang, 2018) and secondly, to identify the model that produces the most accurate results.

The city has experienced elevated levels of air pollutants over the last decade, including concentrations of pm2.5. PM2.5 is highly associated with respiratory ailments and lung cancer. Poor air quality has prompted industries to move to ‘eco-friendly’ production. However, many companies have yet to contribute. This research seeks to help in the prospect of more accurate predictions of the immediate future air quality predictions to enable businesses reduce their polluting activities when poor air quality is expected. Excessive numbers of associated deaths due to low air quality prompted Lary, Lary, & Sattler (2015) to use ML to forecast the pm2.5 concentrations. Consequently, to improve the health of people, proper information-oriented research has begun to increase life expectancy. A research conducted by (Hao, Wang, Shen, Li, & Hu, 2007) focused on the highest degraders of air quality i.e. cement and ceramic factories, thermal power plants and steel plants. Any improved predictions that could be available to these industries could have a significant impact.

1.1 Research Questions

RQ: "*How can we upgrade the short-term air quality predictions for the city of Guangzhou by improving the models to predict the future concentrations of the significant pollutant pm2.5 combined with relevant meteorological data using the deep learning algorithms of LSTM, GRU, CNN and MLP*?"

1.1.1 Sub Question

Sub RQ: *"At what extent, the various meteorological variables (wind speed, temperature, humidity and pressure) are correlated to each other and with the dependent variable pm2.5?"*

1.2 Research Objectives

The research objectives of this exercise are:

Objective 1. A critique review of literature on prediction of air pollution.

Objective 2. To discover the correlation between dependent factors, particularly meteorological components and concentration level of pm2.5 that reduce air quality

Objective 3. Proposed Implementation, Evaluation and Outcome of the MLP, CNN, GRU and LSTM.

Objective 4. A brief comparison of all the implemented deep learning algorithms.

2. Literature Review

2.1. Introduction to Literature

Air quality is critical to human survival and is depreciating in many cities across the globe. Predicting the level and speed of that depreciation has been and is challenging. A better understanding of the data and improved predictions will help planners and scientists improve the outcomes for all. There is the opportunity to use various predictive methods and modelling to have a better knowledge on how air quality degrading over time and the potential impact of other reasons man-made or natural.

Much of the research up to now has focused on time series-based predictive models to help inform how air quality may decline. Additionally, the impact of industrial pollutants has been a target of research and many would advocate curtailing industrial output by individual producers. However, enough scaling back of production is not yet an option for many of these businesses in the context of their financial goals and their long-term survival. For many, a progressive move towards sustainable development would be a more acceptable approach.

Whilst this time series-based forecasting modelling has been beneficial in better perception and predicting when future air quality changes might occur, any improvement in such predictions could add significant value. In relation to the predicting time-based dependencies, there were several developments have been done recently, by utilizing Neural Networks that includes RNN, ANN and CNN.

This segment provides a review of the research to date and the methodologies used to improve air quality prediction, initially with a quick representation of how Neural Networks work and presenting a comparison of the models.

2.2. Examining pm2.5

It has been estimated that the effects of poor air quality on respiratory health is responsible of 6.5 million deaths globally per annum (Priddle, 2016). Air quality is reduced by the presence of tiny particles in the air known as particulate matters (PM). Particulate matters can be present in air in different sized particles but, of all the major air pollutant particles, those known as particulate matters 2.5 (pm2.5), i.e. particles smaller than 2.5 micrometres in diameter are the most toxic.

As far back as 2004, Castanas & Kampa identified that the pm2.5 air pollutants are more life threating to any other particle that present in the air. The researcher Castanas determined that the human body has a better ability to filter out larger particles, whereas pm2.5, because of their tiny size, can enter the lungs causing respiratory illness and premature death.

Therefore, in the fight against air pollution, research into pm2.5 concentrations and predicting the future concentrations is important to estimate future air quality and, ultimately, to save lives.

2.3. Critique Review of Existing Methods and Techniques

2.3.1. Convolutional Neural Networks (CNN)

According to a study conducted by(Huang & Kuo, 2018) where the utilization of CNN was suggested to predict the air quality research

The experiment was able to come up with the two betterment.

1. The analysis had a big impact on the CNN resulting from inclusion of LSTM.

2. the research centered around on the highest producers or air pollution in the city and also put attention on the important details

An experiment done by Chiou-Jye Huang, established that the contribution to urban air quality degradation by the major sources of pollution are:

- vehicle emissions 22%
- coal combustion 17%
- industrial discharge 16%
- other sources 25%, e.g. pollution cried in the air from other countries

In effect, this implies that over 55% of the contributors to pollution could be controlled to reduce or nullify the harmful impact by interventions such as reduction, treatment or elimination.
Unfortunately, because of presence of large numbers of different patterns, it may creates an overfit problem. To address this issue, (Xie, Wang, Wei, Wang, & Tian, 2016) suggested the use of regularization method.

In its conclusion, this experiment states that for forecasting of pm2.5 concentration level, Convolutional Neutral Networks are best option. however, its accuracy is negatively impacted because of some vital deficiencies. CNN could be a best option to challenge the pollution problem only when it gets the strength to overcome the deficiencies.

2.3.2. Artificial Neural Networks (ANN) and Fuzzy Logic

Mishra and his other teammates implemented a totally different methodology, i.e. the utilization of Artificial Neural Networks NN and also Fuzzy Logic for predicting the air quality in New Delhi, India (Mishra, Goyal, & Upadhyay 2015). This research used industrial emissions data exclusively from the geographic state of Delhi. Their research was based on two highly measured geographical districts and highly busy locations, ITO (which is called the hub of industrial manufacturing in Delhi) and the second one was Indira Gandhi International (IGI) Airport. Reading are routinely taken at both these locations. These two districts are geographically well dispersed at two almost opposite poles of Delhi, making these two data sets sufficiently diverse to make them significant.

Their approach took a 3-year data set of daily pm2.5 measurements taken at each location and they have utilized the ANN because it performs outstanding in pattern classification resulting in a self-training model (Gardner & Dorling, 1999). In this same study, neuro-fuzzy logic was another interesting approach adapted (Mishra et al., 2015). The data was collectively combined deriving a fuzzy relationship before analysis in neural networks. The advantage was that if a certain pattern match was present causing a physical polluting haze in a determined time frame, it wasn't required for an exact pattern match to recur to predict a future haze occurring. Instead, a fuzzy algorithm was used to help predict likely recurrences based on near match concentrations and to predict near future (next few days) occurrences considering any pre-emptive actions taken designed to reduce pm2.5 concentrations.

2.3.3. Recurrent Neural Networks (RNN)

Recurrent Neural Networks has been utilized in different regions i.e. modelling of language, image recognition and captioning. In most of the events, it is used to transform the data that is related to time that operates in time-based way. RNN always consider time and sequences for the temporal dimensions. (Athira, Geetha, Vinayakumar, & Soman, 2018). It is used to identify the different types patterns for example text, different spoken words, numerical times data that comes from the stock markets and huge government companies. RNN also applied to different types of images, so that can further decompose into patches and sequences. According to a research conducted by (Tsai, Zeng, & Chang, 2018) it performs well in dealing with the dependencies that occurring from a long time which makes RNN more reliable to use in the research related to time series. As the time intervals increase, the reliability of RNN decreases.

2.3.4. Time Series Analysis

According to the Research done by (Shen, 2012), states that in some big Chinese towns where the levels of airborne pm2.5 exceeds safe levels on over 50 days per year. If it were possible to accurately predict the individual days that safe levels were expected to be exceeded the population could take precautions to avoid or reduce inhalation of these toxic particles. Jaiming considered using different attributes as determinants;

1. values prediction, where the prediction is made based on the level of pm2.5 concentration in the air
2. hidden factor prediction, and following trend prediction, that focuses mainly on a forecasting algorithm for attributes that are indirectly obtained by the information.

Forecasting of timeseries is established and well understood as a predictive methodology especially in the earlier stages of pm2.5 research. However, changes in the number and predictability of influencing factors, has served to reduce the accuracy. A significant example of this is that of unseen variables for example, unpredicted increase in industries., whilst this is categorised as a forecasted event its stretch is not known, and this undermines the resulting predictions.

2.3.5. Long Short-Term Memory (LSTM)

LSTM was created to overcome the problems of RNN in relation to memory-based dependencies. The problem with RNN was getting slow down and make entire process for the machine complicated which made LSTM more demanding. LSTM is very easy to execute as compared to other machine learning or deep learning algorithms. The dependencies issues in RNN can be resolved by using Long Short-term Memory network. A study conducted by (Li et al., 2017) shows that LSTM works as upgraded version of RNN in regard to dependencies. LSTM works as processor as it has the ability to perform read and write operation and remove information from the data. Any comparison to old data inputs could significantly compromise the results as the sources of pollution and the severity of impact change rapidly

2.3.6. Multi-Source Data

A key determinant of the quality of any research is the number of methodologies deployed in conducting the research, i.e. research that uses all possible methods will add greater value to the outcome over research that does not. Such a technique was utilized by (Ni, Huang, & Du, 2017), where the focus entrusts on basic analysis for pm2.5 concentration dependent on data from disparate sources. The intention is to gather data from the maximum number of sources and to obtain a comprehensive prediction of air quality. The research gathered all the required air quality data from the air quality measurements available on the Internet using published air quality measures, data feeds and Internet blogs utilizing some simple web scrapping algorithms. The gathered data was written to a database and the pm2.5 particles calculated. And remaining factors were then evaluated separately before all the data was run through multivariate statistical analysis and a few approaches of neural networks.

2.3.7. Gated Recurrent Units (GRU)

According to (Dey & Salem,2017), Gated recurrent unit (GRU) is an upgraded version of RNN because it has the ability to solve the problem of vanishing gradient. Unlike Long Short-term Memory network, GRU has update gate and reset gate. One of the good things of the GRU is that it can be trained to hold the information regardless of how old the data is without even eradicating the information which is useless for the prediction. By utilizing the both reset and update gate, it can filter out the information as well as it can store the data for a long time. If GRU are trained in well manner so they can come up with the better results in complicated scenarios.

2.4 Examining the issues and drawbacks

A short comparison of all the existing method and techniques that has discussed in the literature review has been done. All the advantages and disadvantages are discussed properly. A comparison table has been attached in the appendix.

3. Research Methodology and Design Specifications

3.1. Introduction:

For the successful execution of this research analysis, we have divided our methodology section into two parts. First part is our Business Model i.e. CRISP-DM because according to business requirements, we can modify it (Brockwell and Davis; 2016). The second part includes the design architecture that has focused on the important components in the research. To get the best results from our dataset, we focused on obtaining a full understanding of the entire process from the very start and to present in a matter that would facilitate reproducing it using comparable data.

3.2. CRISP-DM

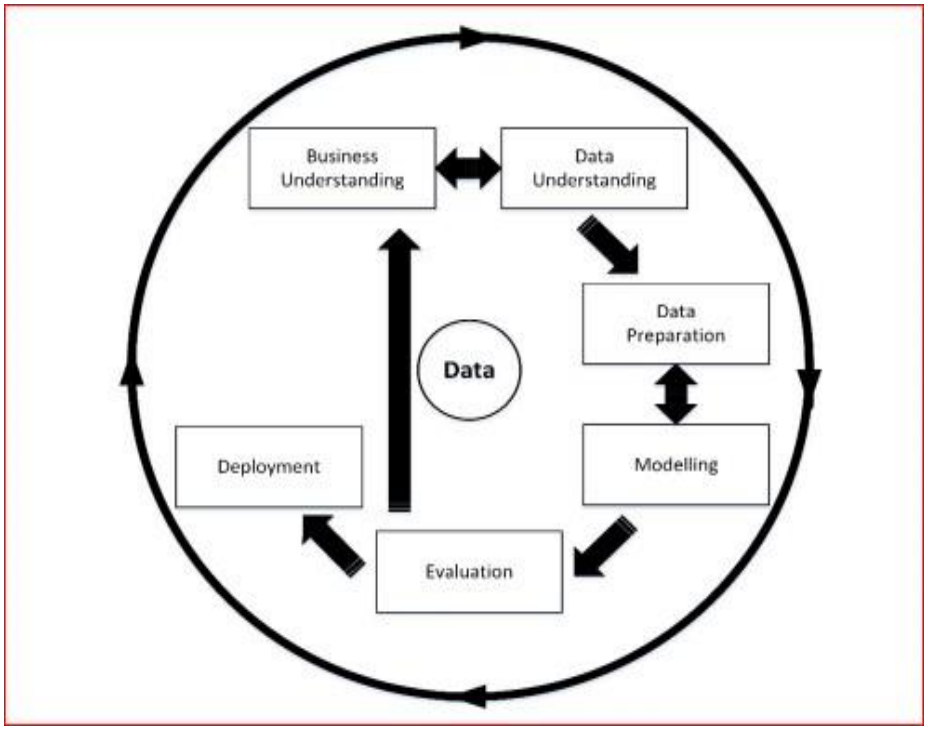


Figure 1: CRISP-DM

CRISP-DM has four levels of breakdown and follows the hierarchical process. (Ncr et al. (2000). When all the prerequisites are set for improving the prediction of air quality degradation in Guangzhou city, it was established there was significant potential to increase the life expectancy and wellbeing of people in the region. In the second phase, the challenge is to discover, after careful research, which variables are necessary to get the best outcome. This phase does not only reduce the complexity but also reduced the size of the data and consequently, decreasing the processing time. After the completion of this research, the data was further processed to edit the data to make it more suited to our research objectives and prepare the interface by removing all the null and missing values which are problematic for the analysis.

This process comes after an thoroughly assessment that centered around forecasting the pm2.5 concentration for various time intervals dependent on the information collected between 2010 to 2014. This would enable an evaluation to approve the accuracy of the prediction models and determine which model's performance is best.

3.3. Design Architecture

The design architecture is divided into two layers. The first layer is the Tier 1 layer, known as Presentation layer and the Tier 2 layer is known as Logical layer.

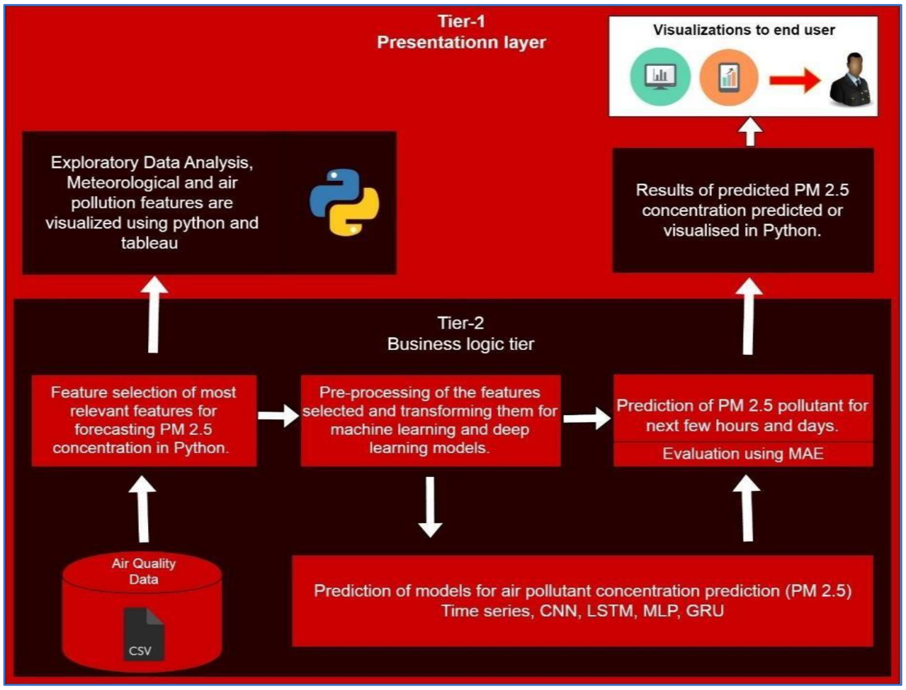


Figure 2: Design architecture

The logical layer is called the backbone of the entire architecture as it placed all the critical aspects of the research in one place. Firstly, the air quality data has been considered from the file for the research to be performed and then initiated for the feature selection and data pre-processing. Afterwards, it leads to Exploratory Data Analysis (EDA) in the Tier 1 layer and the other forecasting algorithms that will predict the air pollutant concentration in the Tier 2 layer. After obtaining the results, the presentation layer has been used to visualise the predicted value against the actual values. Also, Tableau has been used to perform EDA at the presentation layer. The presentation layer has a direct connectivity with the logical layer that allows all logical and modelling permutations to be implemented.

4. Proposed Implementation and Evaluation

4.1. Dataset Description:

The dataset for this exploration has been extracted from the Kaggle where the data was given for the five big cities of China including Guangzhou. The time period of the dataset is from 1st January 2010 to 31st December 2015. In this dataset, the data is updated on an hourly basis and includes meteorological data such as temperature and humidity. The important attributes of the dataset are year, month, day, hour, pm2.5, temp, humi (humidity), pres (pressure), precipitation and lws
(cumulative wind speed), dewp, (dew point). The dependent variable is pm2.5 that contains the pm2.5 concentration level information and the rest of the attributes are used as independent

variables which also portrays a crucial role in predicting the air quality in Guangzhou city. The below figure has shown the variable view of the dataset before any transformation has been applied.

1	No	year	month	day	hour	pm2.5	DEWP	HUMI	PRES	TEMP	cbwd	Iws	precipitat	Iprec
2	1	2010	1	1	0	NA	9.4	76	1015.1	13.5	NW	0.8	0	0
3	2	2010	1	1	1	NA	10.2	83	1015.2	13	cv	0.5	0	0
4	3	2010	1	1	2	NA	10.4	87	1015	12.5	NW	0.6	0.3	0.3
5	4	2010	1	1	3	NA	10.2	89	1014.9	12	NW	1.4	0.6	0.9
6	5	2010	1	1	4	NA	10.4	91	1014.6	11.8	NE	0.6	0.7	1.6
7	6	2010	1	1	5	NA	10.3	91	1014.4	11.7	NW	1.4	0.3	1.9
8	7	2010	1	1	6	NA	10	90	1014.8	11.6	NW	2.5	0	0
9	8	2010	1	1	7	NA	10.4	91	1015.3	11.8	NW	3.7	0	0
10	9	2010	1	1	8	NA	10.3	89	1015.3	12.1	NE	1.2	0	0
11	10	2010	1	1	9	NA	10.5	86	1015.9	12.8	NW	1.2	0	0
12	11	2010	1	1	10	NA	10.9	83	1016	13.7	NE	0.9	0	0
13	12	2010	1	1	11	NA	10.8	82	1016	13.8	SE	0.6	0	0
14	13	2010	1	1	12	NA	10.7	78	1015.3	14.5	NE	0.9	0	0
15	14	2010	1	1	13	NA	10.9	75	1013.8	15.3	NE	2	0	0
16	15	2010	1	1	14	NA	10.8	74	1012.3	15.4	NE	3.3	0	0
17	16	2010	1	1	15	NA	10.8	71	1011.6	16.1	NE	4.6	0	0
18	17	2010	1	1	16	NA	11	73	1011.8	15.8	NE	5.3	0	0
19	18	2010	1	1	17	NA	11.5	77	1012.2	15.5	NW	0.8	0	0
20	19	2010	1	1	18	NA	11.8	81	1012.2	15	cv	0	0	0
21	20	2010	1	1	19	NA	12.1	84	1012.4	14.8	NE	0.7	0	0
22	21	2010	1	1	20	NA	12	83	1012.9	14.9	NE	1.6	0	0

Figure 3: Variable View of the Dataset

4.2. Data Pre-processing:

Data pre-processing is a vital initial step in any machine learning research activity where the use of data is mandatory. To make our data accessible for the research, pre-processing is needed.. . pre-processing is always required when we use neural networks because they use many different layers in their execution. Initially, the data must be made more efficient for the machine learning models by removing the unwanted columns. The objective here is to make the process simple by decreasing the processing time and also to lessen the number of attributes. To produce the correlation, we have combined the three attributes and make it one by utilizing the datetime function. As the below figure shows the combination of time with other variables like year, date and month.

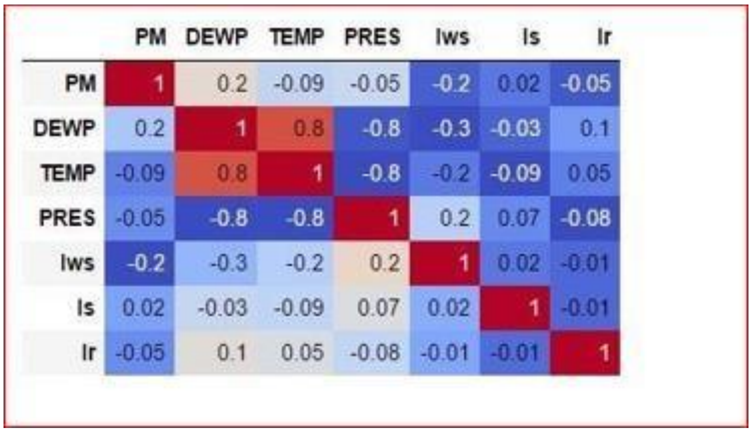
	No	year	month	...	Iprec	datetime	scaled_pm2.5
0	16550	2011	11	...	0.0	2011-11-21 13:00:00	0.089524
1	16551	2011	11	...	0.0	2011-11-21 14:00:00	0.118095
2	16552	2011	11	...	0.0	2011-11-21 15:00:00	0.133333
3	16553	2011	11	...	0.0	2011-11-21 16:00:00	0.148571
4	16554	2011	11	...	0.0	2011-11-21 17:00:00	0.146667

Figure 4: Dataset After pre-processing

4.3. Feature Selection

After the successful pre-processing of data, next step is to prepare the data for feature selection.To concentrate on practicality, feature selection was implemented manually, the research concentrates on air pollution prediction with the help of several meteorological attributes with dependent variable pm2.5. To achieve this and for finding the dependency on each variable with pm2.5, a heatmap graph has presented shows the correlation with all other attributes that have used in the analysis.

The below diagram shows the remaining factors after the less important ones have been removed manually and process of selecting features completed.



	PM	DEWP	TEMP	PRES	Iws	Is	Ir
PM	1	0.2	-0.09	-0.05	-0.2	0.02	-0.05
DEWP	0.2	1	0.8	-0.8	-0.3	-0.03	0.1
TEMP	-0.09	0.8	1	-0.8	-0.2	-0.09	0.05
PRES	-0.05	-0.8	-0.8	1	0.2	0.07	-0.08
Iws	-0.2	-0.3	-0.2	0.2	1	0.02	-0.01
Is	0.02	-0.03	-0.09	0.07	0.02	1	-0.01
Ir	-0.05	0.1	0.05	-0.08	-0.01	-0.01	1

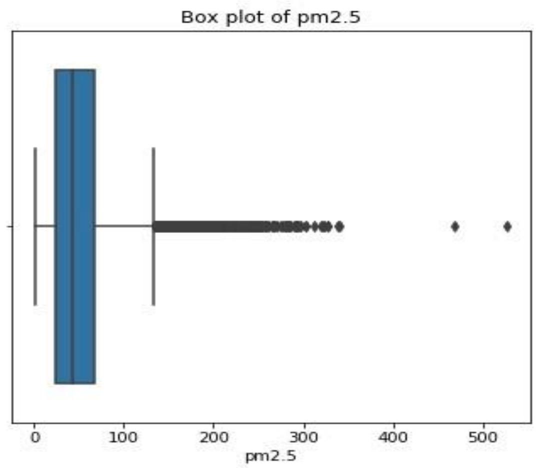
Figure 5: Heatmap showing correlation

This figure clearly attests that the attributes which are close to value 0 and 1 are having a high correlation with the dependent variables and remaining attributes which are not equal to both these values are having a less correlation. So, machine learning and deep learning models can reasonably be implemented in pursuit of improving air pollutants concentration prediction.

4.4. Outliers Checking

Next, to assist in identifying the outliers present in the dataset because with outlier present in the dataset can lead to many errors. Consequently, that can bring down the accuracy of the implemented models.

The following diagrams in figures 6 and 7 shows that some outliers are present,in the dataset.



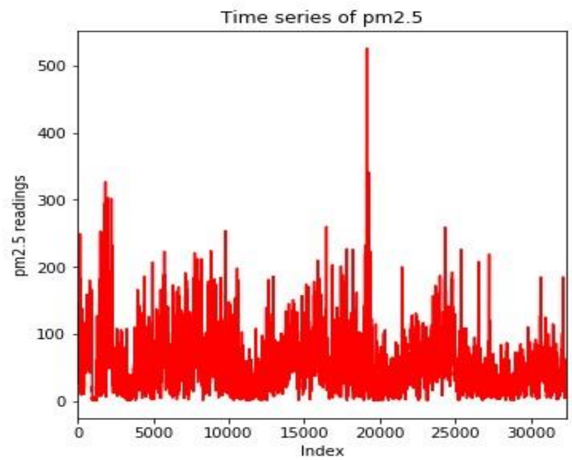
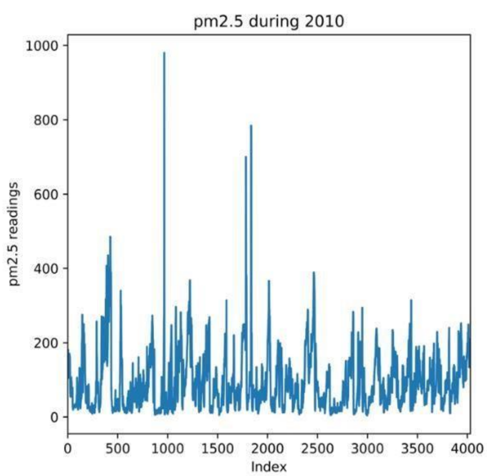


Figure 6: Box plot for PM2.5 concentration

Figure 7: Overall distribution of PM2.5

4.5. Linearity distribution Analysis:

In order to identify whether the data is overly consistent a check of linearity is performed. This is required because inconsistency in data in not good and can create different type of data problems like underfit or overfitting. The data distribution of pm2.5 is depicted in figures 8 and 9. When the data is considered in its entirety it can be considered that no linearity is present. The below figures show the distribution for year 2010 and for the month of January in 2010.



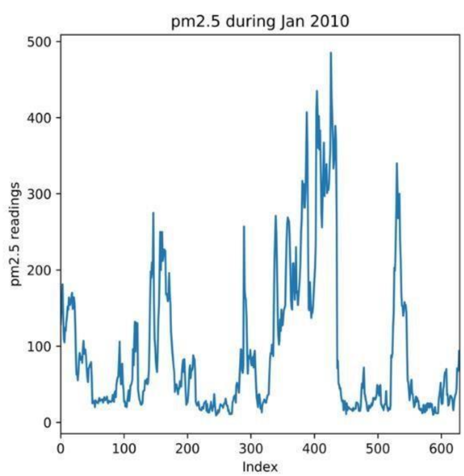


Figure 8: Linearity analysis during 2010

Figure 9: Linearity analysis for January,2010

4.6. Scaling of Data:

In neural network models scaling cane be done by different methods. Normalizing the data is a crucial step that comes under pre-processing, but we have processed it separately to delete the ambiguity from the that we are using according to our requirements. We have divided our dataset into two sets. First the data for training our model comprises of 80% of the data and rest of the data used for the validate the accuracy.

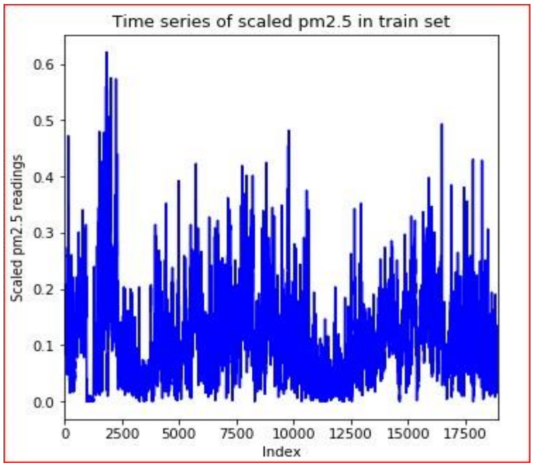


Figure 10: Time series of scale training set

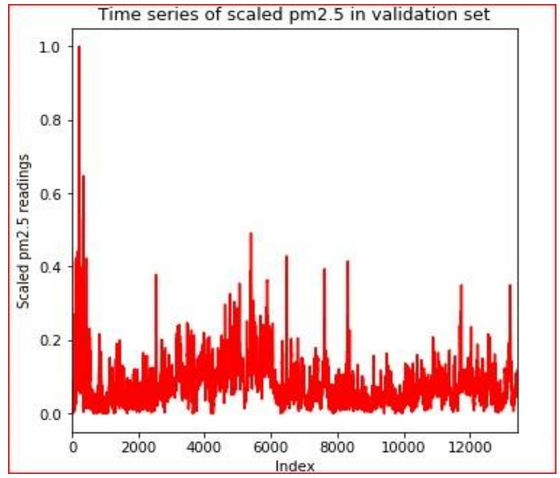


Figure 11: Time series of scale validation

4.7. Libraries, Formulae and Software Used

For the successful execution of this research Python has been used by launching Spyder (Python 3.7) Notebook in Anaconda Navigator (Anaconda3). Excel and Tableau has been used for the visualisation in the research because of their adaptability in finding the correlation in the dataset. Several Spyder Python (3.7) libraries has been used in this project such as seaborn, matplotlib, pandas, datetime, keras, os and sklearn.

Min Absolute Error (MAE) calculations are used in the results evaluation because of the outliers present in the dataset used for this research and there is high risk of deviation and fluctuation in the outcomes which would be unacceptable (Cort & Kenji, 2005). However, MAE provides precise results if there are outliers present in the dataset as contrasted to other squared deviations.

$$\boldsymbol{Mean\ absolute\ error} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} |\boldsymbol{y_j - y}|$$

Where **n** denoted as number of errors,

$|y_j - y|$ denoted as the absolute error

4.8. Implementation of deep learning algorithms

4.8.1. Convolutional Neural Networks

Convolutional Neural Networks are advanced neural networks which are used to image classification, image clustering by utilising various attributes. CNN has the ability to extract features automatically. Before model training, the dataset is split into two sets, first set is to train the data and second set is to validate the data. For training our model, neural networks are used on the train data and for the evaluation of the model, the validation set is used.

Also, to determine the number of epochs that we have used to train the model, Tensor flow backend and Keras Functional API has used. The below figure shows the involvement of all the required layers for implementation of CNN..

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 7, 1)              0
_________________________________________________________________
zero_padding1d_1 (ZeroPaddin (None, 9, 1)              0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 7, 64)             256
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 5, 32)             6176
_________________________________________________________________
average_pooling1d_1 (Average (None, 3, 32)             0
_________________________________________________________________
flatten_1 (Flatten)          (None, 96)                0
_________________________________________________________________
dropout_1 (Dropout)          (None, 96)                0
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 97
=================================================================
Total params: 6,529
Trainable params: 6,529
Non-trainable params: 0
```

Figure 12: CNN parameters output

For adding zero in the starting and ending of all series, Zero padding layer has been used. This layer also ensures that, it does not make any difference in dimension of all the output sequences that occurred.

The Conv1 D layer has three arguments first one indicates the number of features, second argument determined the length of the convolution window and the last argument represents that number of places to shift to the window, called strides. Average pooling layer is used to generate the moving averages of three times unit on the rolling window. The flatten layer is used to reshape all the input and pass it to the output layer.

4.8.2. Long Short-term Memory

LSTM represents an improvement over RNN in that it facilitates three gates to read write and update any data based on the needs utilising memory-based functionality. As a result, LSTM is a good choice where memory-based operations are involved.

The LSTM layers are defined for seven timesteps, here we are using two LSTM layers. The first LSTM returns the output from every seven timesteps in the form of a sequence which is passed to the second LSTM layer which returns output from the last step only. There are sixty-four hidden neurons in each timestep in the first LSTM and therefore the sequence from the first LSTM presents sixty-four features.

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	(None, 7, 1)	0
lstm_1 (LSTM)	(None, 7, 64)	16896
lstm_2 (LSTM)	(None, 32)	12416
dropout_2 (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 1)	33

Total params: 29,345
Trainable params: 29,345
Non-trainable params: 0

Figure 13: LSTM parameters output

4.8.3. Multilayer Perception

Layer (type)	Output Shape	Param #
input_3 (InputLayer)	(None, 7)	0
dense_5 (Dense)	(None, 32)	256
dense_6 (Dense)	(None, 16)	528
dense_7 (Dense)	(None, 16)	272
dropout_3 (Dropout)	(None, 16)	0
dense_8 (Dense)	(None, 1)	17

Total params: 1,073
Trainable params: 1,073
Non-trainable params: 0

Figure 14: MLP parameters output

Multilayer pereptron uses supervised learning method, called back propagation. back propagation used for training purposes. Each and every node present in the MLP used a non-linear activation function because of this function presence and its multiple layer functionality makes it different from the linear perceptron. The activation determines the output of any particular node where a node has given with any inputs. Sometimes MLP are called as “Vanilla” neural networks. As there are three layers used in the figure. Only with input 7 and 16 are having no parameter values. The first layer has received as seven input and

The following layer at that point suggests another sixteen and thirty-two components which combines the past time and generated total 1073 parameters.

4.8.4. Gated Recurrent Units

Gated Recurrent Units are known for their gating mechanism and a better class of Recurrent Neural Networks. GRU acts as an intermediate checkpoint between the Recurrent Neural Networks and Long Short-term memory methods to check the operations for the issued experienced.

As it cane be seen in the figure. The input layer with seven input generating zero number of parameters. The gru_1 and gru_2 are generating the 12672 and 9312 parameters respectively. In total 22017 number of parameters generated by the layers of Gated Recurrent Units.

Layer (type)	Output Shape	Param #
input_6 (InputLayer)	(None, 7, 1)	0
gru_1 (GRU)	(None, 7, 64)	12672
gru_2 (GRU)	(None, 32)	9312
dropout_6 (Dropout)	(None, 32)	0
dense_17 (Dense)	(None, 1)	33

Total params: 22,017
Trainable params: 22,017
Non-trainable params: 0

Figue 15: GRU parameters output

4.9. Evaluation and Results of CNN, GRU, LSTM and MLP

4.9.1. Convolutional Neural Networks

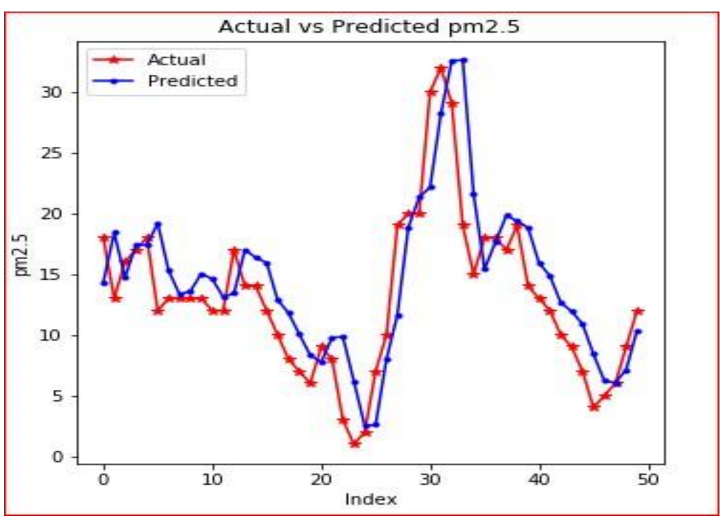


Figure 16: Observed behavior of CNN

In the above diagram, the red line shows the actual behaviour and blue line shows the predicted by prediction models by utilising the processed data. It may be observed, that it isn't fundamentally different in relation to the predicted one. As both the lines follow the same trend, but the predicted line is somewhere up at all points.

The Minimum Absolute Error achieved by the CNN is 6.506.

4.9.2. Long Short-term Memory

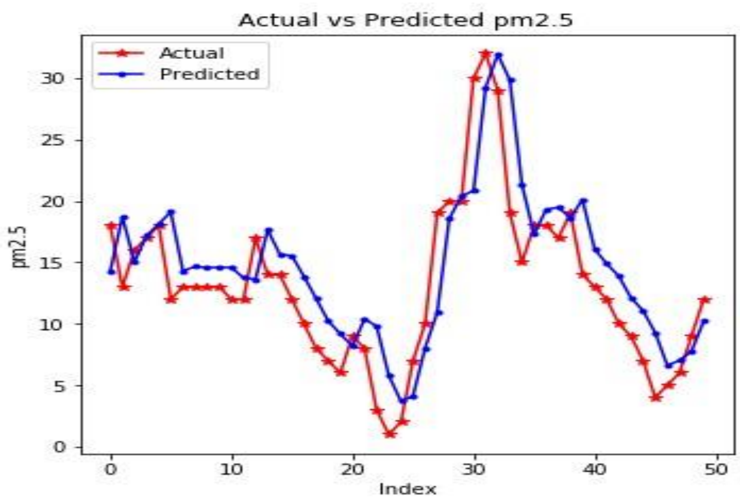


Figure 17: Observed behavior of LSTM

In the above diagram, slight disparities in the forecasted behavior can be noticed. However, the general total remains unchanged when contrasted with the concentration of the actual behavior. The slight inconsistencies could be because of anomalies present in the data, they were not removed by the better version of LSTM.

Hence, the Minimum Absolute Error calculation of LSTM is 6.1854. LSTM performance is better than the CNN.

4.9.3. Multi-Layer Perception

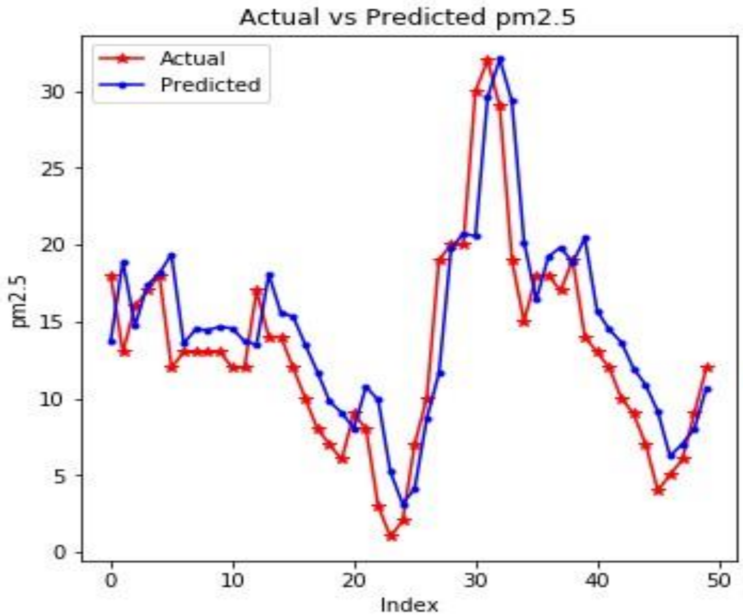


Figure 18: Observed behavior graph of MLP

The regularization strategy actualized is dropout, which suggests the decrease of data problems because of inconsequential unseen layers. In the above figure, there are no big noticeable changes occurred and failed to perform better than LSTM and GRU

The Minimum Absolute Error achieved by the Multilayer Perceptron is 6.2149.

4.9.4. Gated Recurrent Units

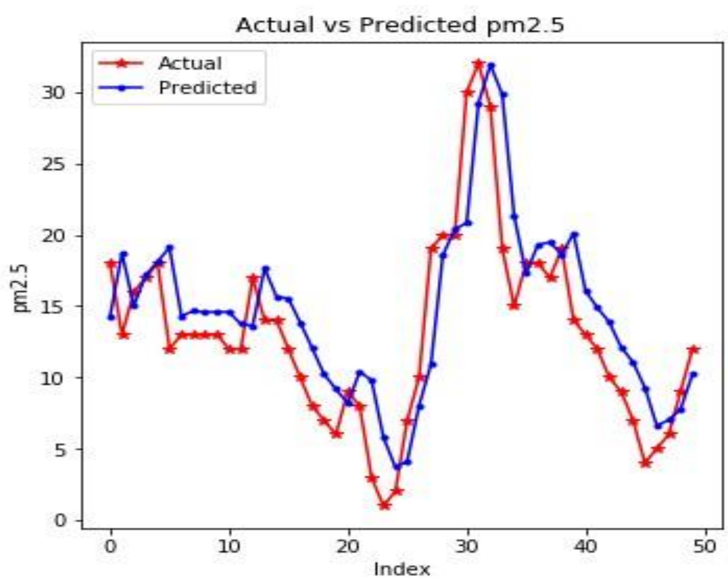


Figure 19: Observed behavior graph of GRU

In the starting, the model was performing well but the model slowly caught down. Hence performed equal as performed by LSTM.

The MAE value achieved by the GRU is equal to the LSTM i.e. 6.185.

5. Results

The results have been computed of all the implemented deep learning algorithms and results achieved are shown in the figure 19.

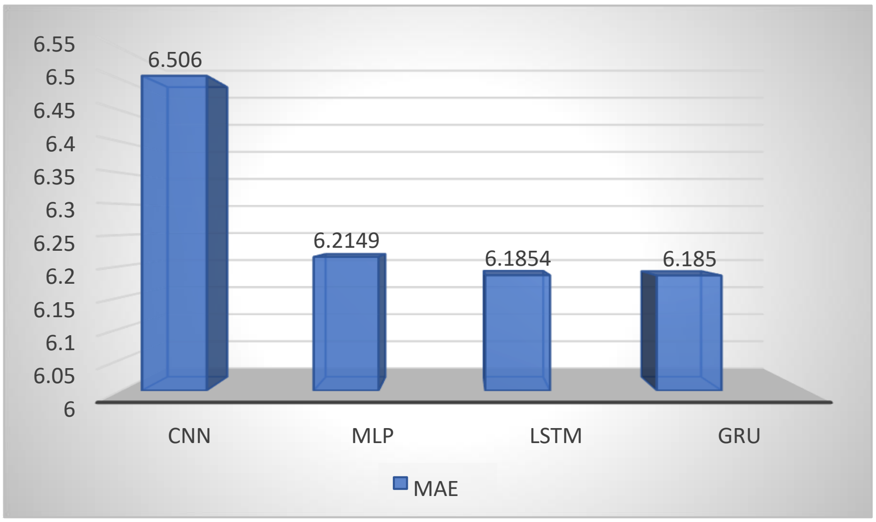


Figure 19: Comparison of models used

6. Discussion

Having executed the models one can assess the results. A universally accepted measure of the accuracy of a prediction model is by reference to the Mean Absolute Error (MAE). Figure 19 is a graphical representation of the MAE results for the models tested. The MAE of the models range from 6.185 (best) and 6.5069 (worst), however, the MAE of GRU and LSTM are very close at 6.185 and 6.1854 respectively.

With its MAE of 6.1854 CNN was the least accurate, a possible explanation for its poor performance might be that it requires a large amount of data due to the large number of parameters.

Of the remaining models MLP is shown to be the next least successful of all the models. This is likely because it is not memory based, this is despite the fact it will have performed better in the hidden layers. The models that performed best were GRU and LSTM. Both these models effectively address a problem inherent in deep learning networks propagated with gradient based methods, that of the 'vanishing or exploding gradients' These models use gating mechanisms to control the flow of long- and short-term dependencies.

LSTM provide the ability to update the data based on the specific research requirements and they further provide the option to remove any outliers if that is desired. Further they provide the ability to update the data feed to meet the project objectives.

In the research output GRU performed best whilst LSTM came a very close second. This can be attributed to the gated mechanisms deployed to improve the quality of predictions. Given this it is reasonable to deduce that, having compared the models implemented in this research project, Question 1 of the research has been answered: GRU and LSTM have shown to improve the quality of predictions and further, of those two models, GRU has been show to perform better in terms of MAE and it also offers a performance advantage.

7. Conclusion and Future work

Having analysed the research output it is reasonable to conclude the research objectives have been achieved in that the predictions of future air quality can be improved and the best model for this is GRU followed closely by LSTM. This will benefit the occupants of cities that suffer from reduced air quality brought about by high concentrations of pm2.5; better accuracy provides residents with the option to take increased precautions when air quantity is expected to be low and provide businesses and regulators with the option to enforce reduced PM2.5 generating activity when airbourne pm2.5 quantities are expected to be high..

A furtherer conclusion is that the results for the Neural Networks models tested the relative performance of each is as broadly as would have been theoretically expected and that GRU and LSTM performed best. However, it should be noted that due to the time available to complete the research the data samples used for learning were relatively small and as a consequence the finding are robust for analysis over shorter durations but it is likely that this will lessen as the duration are extended.

Future research to build on these findings could be tested by further amending the models deployed to attempt better results. Up-sampling could be used to assist training. Further the hyper parameters of the research could be adjusted using LSTM and GRO and review the impact.

In this context further research into improving the prediction of PM2.5 concentration in the air has the potential to increase the accuracy of such predictions and offers the prospect of better future prediction of air quality for the residents, politicians and regulators in areas likely to be impacted.

8.Acknowledgement:

I Virender Singh, student of MSc Data Analytics would like to thanks to my supervisor Dr. Anu Sahni for her continuous help and better supervision throughout the entire research project. Also, I am grateful to my parents and friend who believed in me.

9. REFERENCES:

Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. *Procedia Computer Science*, *132*, 1394– 1403.

URL: https://www.sciencedirect.com/science/article/pii/S1877050918308007

Castanas, E., & Kampa, M. (2004). Human Health effects of air pollution. *Environmental Pollution*, *2*(151), 362–367.

URL: https://www.sciencedirect.com/science/article/pii/S0269749107002849

Cort, J. W., & Kenji, M. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82.

URL:http://climate.geog.udel.edu/~climate/publication_html/Pdf/WM_CR_05.pdf

Dey, R., & Salem, F. M. (2017). Gate-variants of Gated Recurrent Unit (GRU) neural networks. *Midwest Symposium on Circuits and Systems*, 1597–1600.

URl: https://ieeexplore.ieee.org/document/8053243

Fan, S., & Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, *21*(1), 392–401.

URL: https://ieeexplore.ieee.org/document/1583738

Farhani, S., & Ozturk, I. (2015). Causal relationship between CO2emissions, real GDP, energy consumption, financial development, trade openness, and urbanization in Tunisia. *Environmental Science and Pollution Research*, *22*(20), 15663–15676.

URL:https://www.researchgate.net/publication/277410478_Causal_relationship_between_CO2_emissions_real_GDP_energy_consumption_financial_development_trade_openness_and_urbanization_in_Tunisia

Gardner, M. W., & Dorling, S. R. (1999). Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London. *Atmospheric Environment*, *33*(5), 709–719.

URL: https://www.sciencedirect.com/science/article/pii/S1352231098002301

Hao, J., Wang, L., Shen, M., Li, L., & Hu, J. (2007). Air quality impacts of power plant emissions in Beijing. *Environmental Pollution*, ***147*(2), 401–408.**

URL:https://www.researchgate.net/publication/6888516_Air_quality_impacts_of_power_plant_emissions_in_Beijing

Huang, C. J., & Kuo, P. H. (2018). A deep CNN-LSTM model for particulate matter (Pm2.5) forecasting in smart cities. *Sensors*, *18*(7).

URL: https://www.ncbi.nlm.nih.gov/pubmed/29996546

Lary, D. J., Lary, T., & Sattler, B. (2015). Using Machine Learning to Estimate Global PM 2.5 for Environmental Health Studies. *Environmental Health Insights*, *9*(S1), 41–52.

URL: https://www.ncbi.nlm.nih.gov/pubmed/26005352

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

URL: https://www.researchgate.net/publication/277411157_Deep_Learning

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, *231*(September), 997–1004.

URL: https://www.sciencedirect.com/science/article/pii/S0269749117307534

Liu, B. C., Binaykia, A., Chang, P. C., Tiwari, M. K., & Tsao, C. C. (2017). Urban air quality forecasting based on multidimensional collaborative Support Vector Regression (SVR): A case study of BeijingTianjin-Shijiazhuang. *PLoS ONE*, *12*(7), 1–17.

URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179763

Mass, C., & Portman, D. (1989). Major Volcanic Eruption and Climate: A Critical Evaluation. *Journal of Climate*, *2*, 566–593.

URL:https://journals.ametsoc.org/doi/abs/10.1175/15200442%281989%29002<0566%3AMVEACA>2.0.CO%3B2

Mishra, D., Goyal, P., & Upadhyay, A. (2015). Artificial intelligence based approach to forecast pm2.5 during haze episodes: A case study of Delhi, India. *Atmospheric Environment*, *102*, 239–248.

URL:https://www.researchgate.net/publication/268870569_Artificial_Intelligence_Based_Approach_to_Forecast_PM25_during_Haze_Episodes_A_Case_Study_of_Delhi_India

Ni, X. Y., Huang, H., & Du, W. P. (2017). Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data. *Atmospheric Environment*, *150*, 146–161.

URL: https://www.sciencedirect.com/science/article/pii/S1352231016309451

Priddle, R. (2016). World Energy Outlook - Special Report Energy and Air Pollution. In *World Energy Outlook - Special Report*.

URL: https://www.sciencedirect.com/science/article/pii/S1352231016309451

Shen, J. (2012). *PM 2.5 concentration prediction using times series based data mining*.

URL: http://mickeystroller.github.io/resources/DM_Project_Jiaming_Shen.pdf

Tsai, Y. T., Zeng, Y. R., & Chang, Y. S. (2018). Air pollution forecasting using RNN with LSTM. *Proceedings - IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, IEEE 16th International Conference on Pervasive Intelligence and Computing, IEEE 4th International Conference on Big Data Intelligence and Computing and IEEE 3*, 1074–1079.

URL:https://www.researchgate.net/publication/328605377_Air_Pollution_Forecasting_Using_RNN_with_LSTM

Wirth, R., & Hipp, J. (2010). CRISP-DM: Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery* Towards a *and Data Mining*, (24959).

URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.198.5133

Xie, L., Wang, J., Wei, Z., Wang, M., & Tian, Q. (2016). DisturbLabel: Regularizing CNN on the loss layer. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4753–4762.

URL:http://openaccess.thecvf.com/content_cvpr_2016/papers/Xie_DisturbLabel_Regularizing_CNN_CVPR_2016_paper.pdf

Yang, M. (2018). *A Machine Learning Approach to Evaluate Beijing Air Quality*.

URL: https://www.math.ucdavis.edu/files/2015/2717/8083/Mingy_Yang_Spring_2018.pdf

Brockwell, P. J. and Davis, R. A. (2016). Introduction to time series and forecasting, springer.

URL: https://www.springer.com/gp/book/9781475777505

Ncr, P. C., Spss, J. C., Ncr, R. K., Spss, T. K., Daimlerchrysler, T. R., Spss, C. S. and Daimlerchrysler, R. W. (2000). Crisp-dm 1.0.

URL: https://www.the-modeling-agency.com/crisp-dm.pdf

Appendix 1 Some Problems faced in the Research Models

Option	Advantages	Disadvantages	Reference
Artificial Neural Networks with Fuzzy Logic	Neural networks are recommended for Identifying Patterns in the data whist fuzzy logic can assist broaden the range of what can be achieved	There is a decrease in performance 9time to complete) where Fuzzy logic is used, The impact on independencies due to the unverified back end inherent in ANN would need consideration even more so in this research as the algorithm would need frequent (daily) runs.	(Mishra et al., 2015)
Convolutional Neural Networks	Where all the local features are correlated in any new data set then CNN can extract features automatically	Firstly, CNN computations are resource heavy and requires large data available for training. Also, on pattern detection a loss value is generated resulting in a future overfit.	(Huang & Kuo, 2018)**,** (Xie et al., 2016)
Long Short Term Memory	LSTM successfully addresses RNN's long-time dependencies issue	Results might change if historic data needs to be considered, consequently we need to particularly careful of historic date until confirmed	(Li et al., 2017)
Multi-source data	Multiple data sources facilitate broadening the data set which results in better training	Unverified additional data sources might not be reliable and so could compromise the results. In this research 'blogs' were used, these cannot be verified	(Ni et al., 2017)
Recurrent Neural Networks	RNN benefits from a good ability to detect changing patters over time.	There are well recognised issues of exploding and vanishing gradient resulting from increases in the data intervals where longer term analysis is used	(Athira et al., 2018)**,** (LeCun et al., 2015)**,** (Tsai et al., 2018)