***Abstract – Air pollution is a growing threat towards society and various measures are being taken recently in order to avoid it. The problem of concern which remains is the efficient prediction of air pollution in order to work in the right direction for reducing the same. LSTM Neural Network can be used for the same which is feed with the air pollution data. Weather conditions too affect the air pollution, which can be used to further improve the accuracy by processing whether data through Deep neural network and integrating both as a model.***

1. **INTRODUCTION**

The basic human needs for survival are Air, Water and Food. Air being the most important of them. Due to Industrial revolution and Urbanisation the health of environment is highly affected resulting in pollution. Out of these pollutions, air pollution directly affects the health of individuals. The most common elements of air pollution are PM 2.5, PM 10, SO2, NO2, CO, O3, etc [<https://www.health.nsw.gov.au/environment/air/Pages/outdoor-air-pollution.aspx>]. The most harmful amongst being PM (Particulate Matter) 2.5, as their size is less than 2.5 microns and can easily penetrated through filters. Theses particles if continuously inhaled through humans can be very lethal causing various health problems like cardiovascular and respiratory problems, increased stress on heart and lungs, increasing ageing of lungs, shortening life-span etc [<http://www.sparetheair.com/health.cfm>]. The Air Quality Index (AQI) can be used for tracking the health of air of the region [<http://scorecard.goodguide.com/env-releases/def/cap_psi.html>]. The forecasting of AQI can be used as a safely measure to initiate alertness in the region and the necessary action could be taken in order to reduce that.

The growth in technologies like Artificial Intelligence, Machine Learning, Deep Learning over the years can be used in order to design a forecasting model. In this paper we present a neural network which has input parameters like whether API, climate API, along with previous AQI and can forecasts the AQI of the region.

In this paper we present a model which can be implemented for AQI forecasting taking data inputs from the air pollution detection system installed throughout the region.

**2. MODEL**

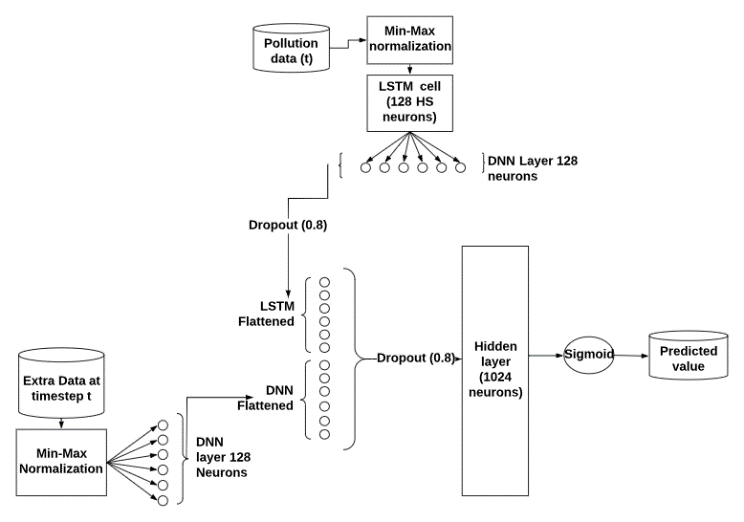
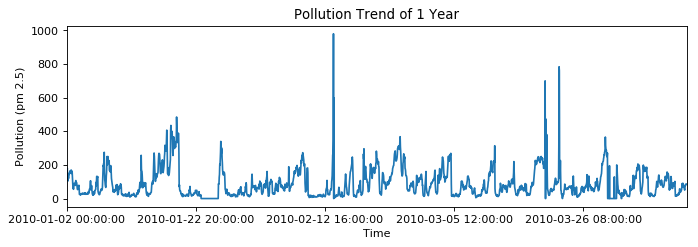
The data fields are separated and fed into 2 different networks. The value at time t of the variable which is to be predicted at time t+1 is passed into an LSTM cell. The LSTM cell has 128 nodes in the hidden state. The other information that affects the target variable is fed into a separate DNN layer with 128 nodes with ReLU activation. Both these layers are concatenated into a single layer. Nodes of the concatenated layer is fed into a hidden layer with 1024 nodes with a ReLU activation function. The output from all the nodes in this layer is fed to a single node with sigmoid activation to get the final output. The output should be scaled inversely to get the true predicted value at time t+1. The flow of the model is depicted in Figure 1. The separate DNN layer helps the model to understand the drastically increasing and decreasing trends better than the conventional LSTM cell can.

Figure 2. Sample of pm 2.5 from the dataset

Figure 1. Flow diagram of the model

**3. DATA**

**3.1 Data Description**

The dataset used was the Beijing PM 2.5 data which has the calculated values from January 1st, 2010 to December 31st, 2014. The fields in the set includes the time, pm 2.5 readings, temperature, wind speed (lws), wind direction (cbwd), pressure, dew, snow (ls) and rain (lr). It is visible in Figure 2 that the variation in the pm 2.5 levels in the dataset is drastic which makes it a good set to test the robustness of the model.

**3.2 Data preparation**

The time steps in the data set was changed from hourly pm 2.5 levels to the daily max level of pm 2.5. Changing the focus of the data set resulted in a better forecasting accuracy. The sections where the data was missing was filled with a value which was randomly selected between the and where is the standard deviation of the column .The data of month and the wind direction were One Hot Encoded with one vector column removed to avoid the dummy variable trap. All the data including the pm2.5 levels were scaled using *min-max normalization*.

The batch generation function returns separate datasets for the DNN and LSTM network. Both contain their respective data from timestamp 0 to t-1. The output dataset for training is the pm 2.5 levels from timestamps 1 to t.

**4. MODEL TRAINING AND TESTING**

**4.1 Training**

The model was trained on batches of 60 days per iteration for 40,000 iterations. Huber loss function was used to optimize the model and was trained using gradient descent with Adam optimizer for the second order behaviour. The initial value of learning rate was 1x10-4 and the hold probability for dropout was initialized to 0.8.

**4.2 Testing**

The model was tested using 2 criteria and was put against a simple LSTM cell network trained on the same data set in a similar manner to compare the results. Both the tests used the *mean absolute error* to measure the accuracy.

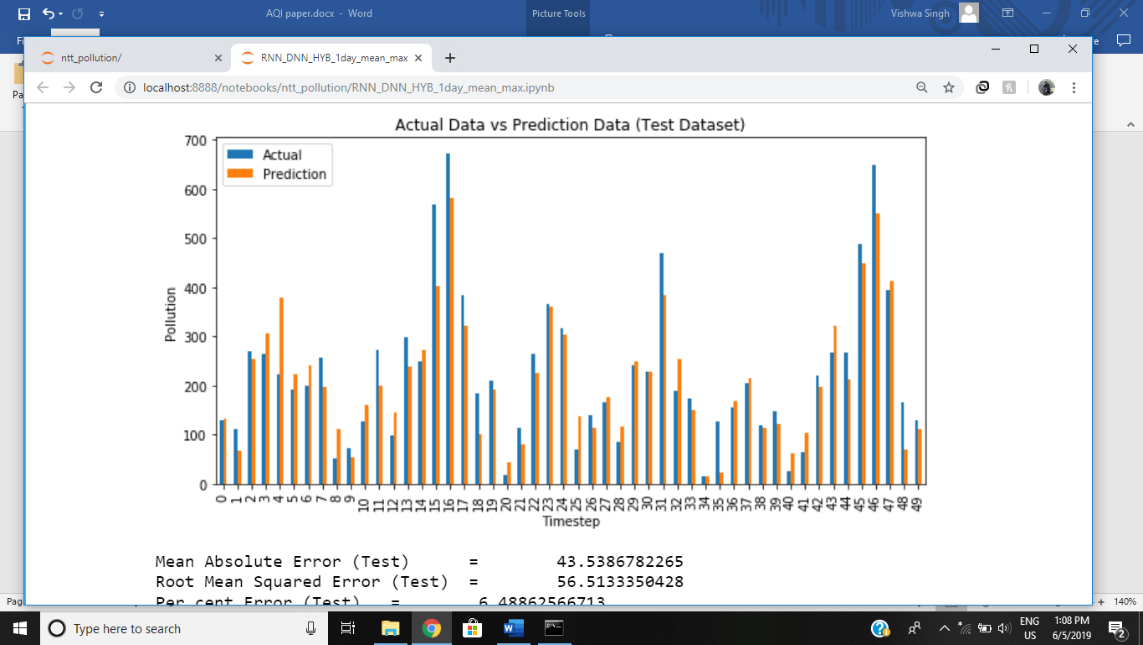
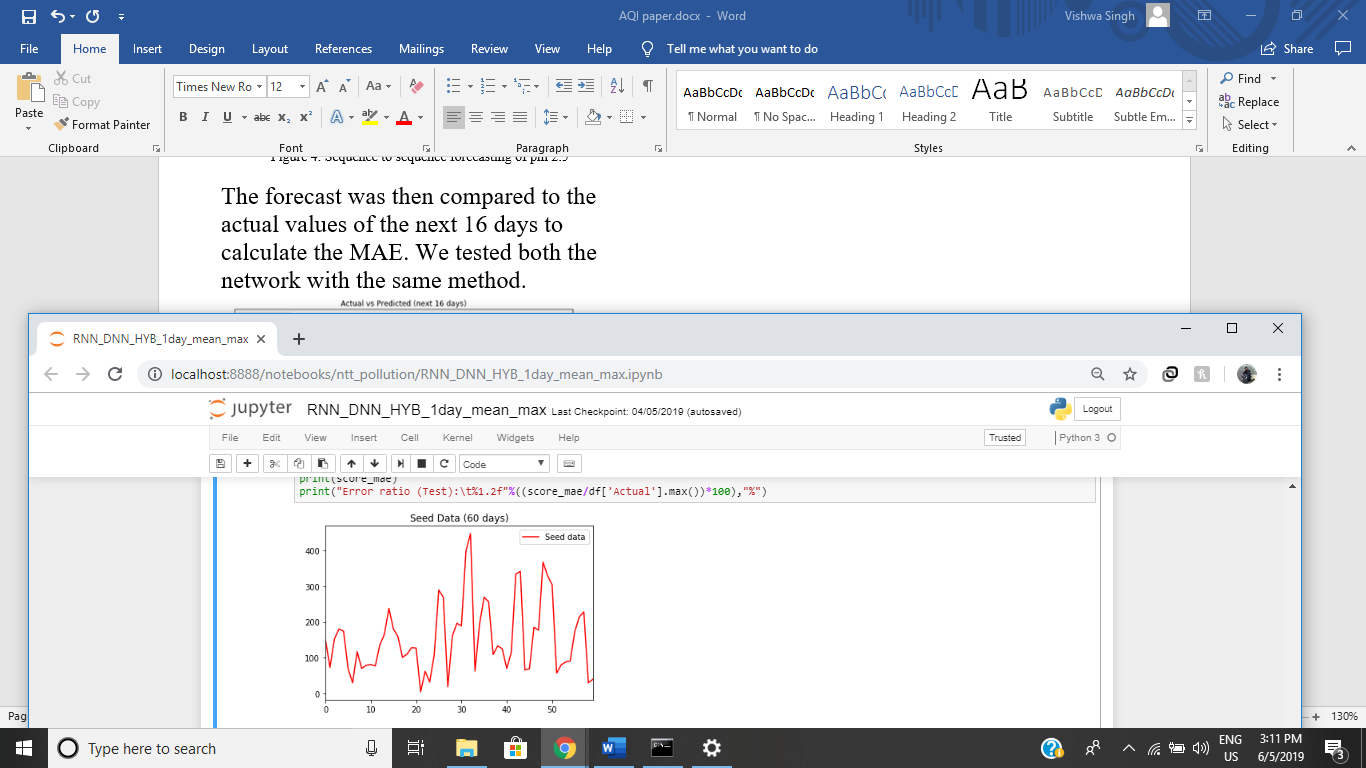
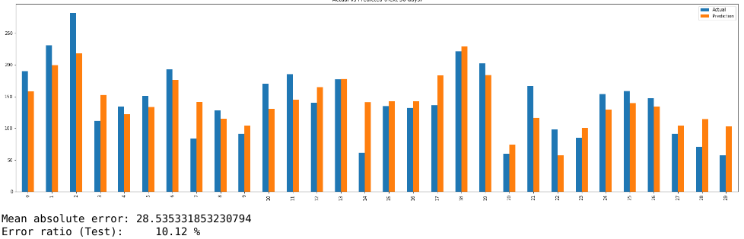
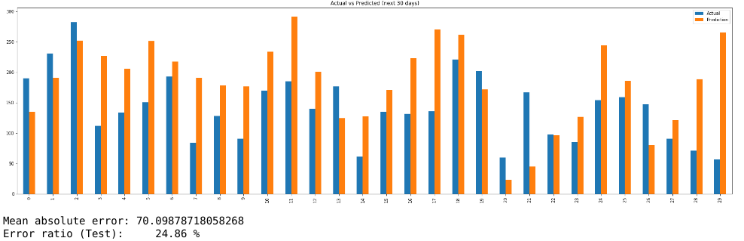
 In the first test, we fed the data from the set for a timestamp t and fetched the prediction for the next timestamp t+1. The prediction was compared with the actual value at that timestamp. This test was done for 50 days starting from different timestamp and calculated the MAE (figure 3).

Figure 3. The prediction vs actual graph of hybrid network

The following table 1 is the result of the test conducted on samples starting from randomly selected points.

Figure 5. Seed and the plot of forecast vs actual pm 2.5 values for 30 days of the hybrid network (top) and LSTM cell (bottom)

Table 1. MAE values for test 1

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Hybrid Model** | **Simple LSTM cell** |
| 1 | 55.644 | 63.370 |
| 2 | 44.401 | 55.208 |
| 3 | 55.514 | 57.919 |
| 4 | 39.696 | 46.755 |
| 5 | 62.459 | 67.179 |

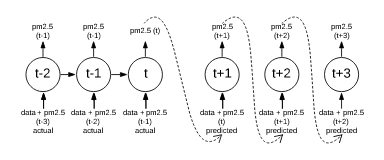
In the next test, we forecasted the pm 2.5 level for the next 16 days and next 30 days. The prediction was done using the sequence to sequence forecasting method (figure 4). We fed a seed of the last 60 days to load the memory

Table 2.1 MAE values for test 2 for 16 days forecast

Figure 4. Sequence to sequence forecasting of pm 2.5

neurons of the network and make the prediction trend more accurate.

The forecast was then compared to the actual values no those days to calculate the MAE (figure 5). We tested both the network with the same method.

The following tables shows the result of the second test for both the network conducted on samples starting from randomly selected points.

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Hybrid Model** | **Simple LSTM cell** |
| 1 | 36.426 | 61.795 |
| 2 | 68.593 | 75.087 |
| 3 | 26.436 | 34.060 |

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Hybrid Model** | **Simple LSTM cell** |
| 1 | 32.783 | 58.634 |
| 2 | 81.193 | 90.360 |
| 3 | 28.535 | 70.098 |

Table 2.2 MAE values for test 2 for 30 days forecast

**5. CONCLUSION**

The results show that the hybrid network performs better than a conventional LSTM cell, specially in the cases when the change is drastic.