PM10 Density Forecast Model Using Long Short Term Memory

Jung-Hwan Park
Dept of Computer Engineering
Sejong Univ, 209 Neungdong-ro
Gwangjin-gu, Seoul, South Korea
+82-2-465-5266
nbicjh@gmail.com

Seong-Joon Yoo*
Dept of Computer Engineering
Sejong Univ, 209 Neungdong-ro,
Gwangjin-gu, Seoul, South Korea
+82-2-3408-3755
sjyoo@sejong.ac.kr

Kyung-Joong Kim
Dept of Computer Engineering
Sejong Univ, 209 Neungdong-ro
Gwangjin-gu, Seoul, South Korea
+82-2-3408-3838
kimkj@sejong.ac.kr

Yeong-Hyeon Gu
Dept of Computer Engineering
Sejong Univ, 209 Neungdong-ro,
Gwangjin-gu, Seoul, South Korea
+82-2-465-5266
yhgu@sejong.ac.kr

Keon-Hoon Lee
Dept of Computer Engineering
Sejong Univ, 209 Neungdong-ro
Gwangjin-gu, Seoul, South Korea
+82-2-465-5266
leekeonhoon@gmail.com

U-Hyon Son
R&D Amway Korea
517, Yeongdong-daero,
Gangnam-gu, Seoul, South Korea
+82-2-557-2566
Uhyon_Son@Amway.com

***Abstract*— This paper suggests a PM10 forecast model using Long Short Term Memory (LSTM). Data used for the study are collected from Seoul, Korea for the period of January 2005 up to March 2016. As the collected data has a lot of noise, the moving average technique is used to preprocess data for smoothing. Time series data of PM10 was converted into 30-day sequence data to use it as the input data for LSTM. LSTM learns through the sliding window process where sequence data moves to the space adjacent to it. The linear regression and recurrent neural network models are compared to evaluate the performance of LSTM. From the result, the suggested model showed a 500% improvement over linear regression and 100% over the recurrent neural network for its performance level.

Keywords: Fine Dust, Long Short Term Memory, Machine Learning, Particulate Matter, Recurrent Neural Network , Sequence data, Time Series

I. INTRODUCTION

This paper suggests a time series data-based PM10 density forecast model using Long Short Term Memory (LSTM) [1]. The data used here are the PM10 data of Seoul, Korea from January 2005 to March 2016, downloaded from the Arikorea [2] website.

PM10 refers to the particulate matter 10 µm or below in diameter that floats in the air for a long period of time. PM10 has an adverse effect on skin and the respiratory system, and is a matter worth being cautious about, according to worldwide climate change. In order to forecast such particulate matter, many forecasting models using statistical techniques and machine learning have been studied. Previous studies on the PM10 forecast were based on the models of linear regression and neural networks [3]. Unlike them, this paper suggests a PM10 forecast model using LSTM.

PM10 data of Seoul, Korea from January 2005 to March 2016 are first collected. Although there is data from other regions available, PM10 data of Seoul are collected as it has the highest population.

PM10 data collected are preprocessed with the Moving Average (MA). Based on the preprocessed data, a learning phase using the LSTM model is conducted and the density of PM10 is forecasted. Comparing the Recurrent Neural Network (RNN) [5], RNN has a drawback of not being able to study the data for long periods of time. LSTM is a model that overcomes this drawback of the RNN.

A 30 days MA value randomly assigned is used to convert the time series data into sequence data. The converted sequence data is used as input data for the LSTM model and the output data is forecast for the next day. Such a process of moving to the immediately adjacent dataset is called *sliding window* and the LSTM model is trained based on this. Rectified Linear Unit (ReLU) [6] is used as the activation function utilized at the output of the model. The performance of the non-stationary model is optimized by using the Root Mean Square Propagation (RMSProp) [7] optimization function. In addition, it is comparatively evaluated against another similar optimization function, Adagrad [8]. In order to measure the performance of the LSTM forecast model, this paper uses Root Mean Square Error (RMSE)[9]. RMSE is used to illustrate the differences between the real value and the value of the trained model when forecasting. The performance of the model improves as the value of RMSE approaches zero. Hence, RMSE is used to compare its own performance with the existing models previously studied. The linear regression model and the RNN of a neural network are used with Adagrad of LSTM for comparison. The comparison analysis shows that the suggested LSTM model displayed a performance 500% better than that of linear regression and 100% better than that of the RNN model.

The content of this paper is as follows: Section 2 analyzes the related research, and Section 3 explains the model used. Section 4 describes the data collection method, whereas Section 5 describes the PM10 forecast method by using LSTM. Section 6 explains the evaluation on the forecast model, and finally, Section 7 concludes this paper as well as describes the direction of future research.

II. RELATED WORK

Section 2 describes the related work on the data and the model used in this paper. Areas of research include PM10, the

* Corresponding author

PM10 forecast using machine learning and the forecast using LSTM. Related researches in these areas are described.

A. Research on PM10

Particulate matter in the air can be classified into TSP, PM10, and PM2.5. PM10 refers to the particulate matter 10 µm or below in diameter that floats in the air for long periods of time.

PM10 increases due to the influence of climatic conditions [10]. PM10 can also increase rapidly due to the influence of smog and yellow dust [11].

PM10 is often regarded as not very harmful to the human body but in actuality, it is harmful to skin and respiratory systems [12]. WHO [13] has established a PM10 guideline and advice accordingly, whereas the United States [14] offers a guideline to protect the health of the people and to foster a clean air environment.

B. Research on PM10 forecast using Machine Learning

PM10 has been forecast in much of the research previously. D. Vlachogiannis [15] used the feed-forward neural network, linear regression, and the hybrid clustering algorithm model to forecast PM10. He used the time series data of Helsinki, Finland from November 31 to December 1, 2003, and used characteristics like PM10, O_3, relative humidity, wind speed, and wind direction. S. Thomas [16] suggested a method using the regression model and the neural network model to develop a model to forecast PM2.5 and carbon dioxide on highways. M. Misity [17] used the mixture linear regression model to forecast time series PM10 data from 2007 to 2011. In total, six different areas were used for the research and linear regression and the EM algorithm were combined and studied. While using the machine learning approach, H. A. Hamid [18] used Artificial Neural Networks (ANN) and Support Vector Machines (SVMs) to forecast the PM10 density of the roadsides of downtown Hong Kong. PM10 data and weather data from 2008 to 2011 were used for the research and PM10 was forecast based on map-based learning using two levels (0 and 1). PM10 forecast research using machine learning has been actively engaged in practical applications [19].

Existing research has used the linear regression model and the neural network model of machine learning to conduct the research. However, there has not been a case where LSTM was used to forecast PM10. Therefore, this paper conducts research by comparing LSTM to linear-regression of the existing machine-learning algorithm, as well as to the RNN of a neural network.

C. Forecast research using LSTM

RNN is a type of neural network and LSTM has been studied as one of the representative RNN. LSTM is a model that does sequential computation by converting time series data into the form of sequence data. Many data like these that represent stocks, health, and network systems include time series data. By converting a part of time series data into the sequence data form and through continual learning, the LSTM model can be developed.

C. Lipton [21] used LSTM to solve the classification problem of medical data. Once the patients' Electronic Health Record (EHR) time series data was classified by LSTM, the medical data was learned and the patients were classified in this research. P. Malhotra [22] studied the ideal value data by using LSTM. Ideal value data is a phenomenon that appears when there is a sudden increase of data in an average data distribution. Hence, this research converted time series data into sequence data, studied the stacked LSTM model, and forecast the future ideal value data. Besides this, there has been research forecasting time series data by using RNN and the LSTM model [23].

Many existing researches use variables to forecast. The problem with this research is that data represents only short periods of time and the forecast accuracy is low. Therefore, this paper uses data spanning 10 years and uses the univariate LSTM model to forecast PM10.

III. MODEL DESCRIPTION

Section 3 explains the model used in this study and its evaluation. The model used in this paper is linear regression, RNN, and LSTM; these models are explained in detail.

A. Linear Regression

Linear regression is a statistical method used in forecasting the relationship between variables. There are different types of linear regression. Simple linear regression only uses a single independent variable to dependent variables, whereas multiple linear regression uses many independent variables. For example, if price change or competition influences the sales of a specific brand, the linear regression model can forecast the impact of each element towards the sales. Eq. (1) is a mathematical formula for linear regression. α is the y-intercept, β is the slope, and ε is the error.

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{1}$$

B. Recurrent Neural Network (RNN)

RNN is used to process sequential information. RNN has a loop that can study data continuously, and this loop helps maintain information. Therefore, it has a structure in which data of the past can influence the future. For example, when a person reads an essay, he or she understands the next word or sentence according to the previous context. This is the same as thinking based on words and sentences already read instead of reading all content in the essay. Just as being able to understand the following content by remembering the context right before, RNN is a model based on such structure. RNN is most commonly used in voice recognition, translation, and image captioning. It can be divided into LSTM, Gated Recurrent Unit (GRU) [24] model, etc.

Eq. (2) is a mathematical formula for RNN. In Eq. (2), $x_t$is input vector; h_t is hidden layer vector; y_t is output vector; W, U and $b$is parameter matrices and vector; and $\sigma_h$and $\sigma_y$is activation functions.

$$s_t = \tanh(Ux_i + Ws_{t-1})$$

$$\sigma_t = \text{softmax}(Vs_t) \quad (2)$$

C. Long Short Term Memory (LSTM)

The strength of RNN is being able to connect the current work with previous information. However, RNN shows problem in connecting with information as the distance from the previous data increases. This is called long-term dependency. Hence, LSTM has been studied to resolve the long-term dependency problem of RNN. The core of LSTM is the continuous cell state coming through the gate, which is the horizontal line passing the upper part of the diagram. The continuous cell state is called the conveyer belt. Information coming through the conveyer belt continues without change. LSTM can add or delete information through the gates of input, forget and output. In conclusion, a gate plays the role that lets information pass through selectively and by using this, learning continues by deleting old data.

Eq. (3) is a mathematical formula of LSTM. LSTM is calculated by using gate vector. To explain gate vector, $f_t$is the *forget gate vector* and acts as the weight of remembering old information. $i_t$is the *input gate vector* and acts as the weight of acquiring new information, whereas $o_t$is the *output gate vector,* playing the role of selecting output candidates. $x_t$is the input vector; $h_t$is the output vector; $c_t$is the cell state vector; and $W, U$and b are parameter matrices and a vector, respectively. $f_t, i_t,$ and $o_t$are gate vectors. Within LSTM, there are 3 activation functions. σ_g is the sigmoid function, σ_c is the hyperbolic tangent, and $\sigma_t$is the hyperbolic tangent, used as $\sigma_h(x) = x$.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (3)[1]$$

IV. DATA SET

This section describes the collection of the data set used in this paper, as well as the preprocessing method. Ideal value data and Not a Number (NaN) data were removed from the collected data and preprocessing was done by using MA, which is a smoothing technique.

Data used for forecasting PM10 are collected from the AirKorea website managed by the Korea Environment Corporation. Data collected are PM10 data of the Seoul region from January 1, 2005 to March 31, 2016. Preprocessing is done to remove NaN data and ideal value data. NaN data occurs when there is an error in the PM10 collection equipment. Ideal value data occurs when PM10 number increases rapidly, and it is removed because it brings the performance of the model down.

Since daily data is used as PM10 data, noise is dense. Noise occurs without certain rules and it can be removed by a smoothing technique using MA. Many different values can be input into MA, but a random 30-day MA is used to remove data noise before using the data. Fig. 1 (a) is the raw data of PM10 and (b) is the result of preprocessing using MA.

When using a 30-day MA, the average data value of N days is used as the data of N+1 days. Hence, the data from day 1 to day 30 are excluded.

V. PM10 FORECAST MODEL USING LSTM

Section 5 describes the formation of the LSTM PM10 forecast model used in this paper. LSTM learns based on sequence data. Therefore, the sequence data used as input data and the parameter used as well as the LSTM model learning method are explained.

A. Generating sequence data

LSTM learns data by using sequence data. Hence, time series data needs to be transformed into sequence data in order to create the feature to use as input data of LSTM. This study restructures data after preprocessing as a single sequence and learns the LSTM model.

Fig. 2 is an example explaining the sequence used in this paper. The first sequence of Fig. 2 is data from *t* day to *t+2* day forecasting *t+3* day, whereas the second sequence uses *t+1* day to *t+3* day to forecast *t+4* day.

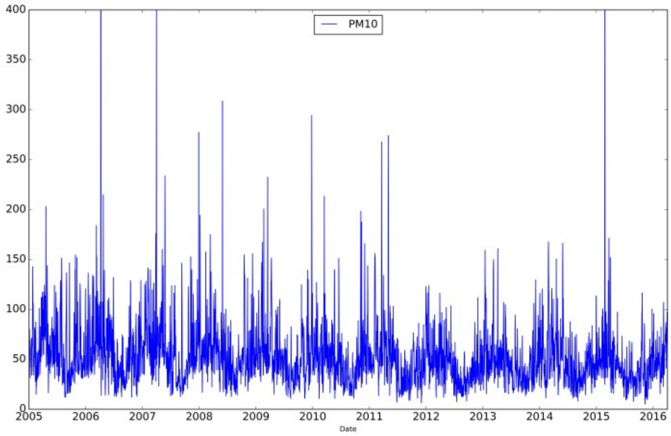
B. PM10 Forecast LSTM model

The overall flow of the PM10 LSTM model is shown in Fig. 3. The LSTM model was developed based on the RNN and is composed of cells with many gates attached. Sequence data enters through the input gate, and then goes through the mini-batch process of partial learning; the forecast value on one of the sequences is output through ReLU among activation functions right after output gate. The LSTM model repeats this process.

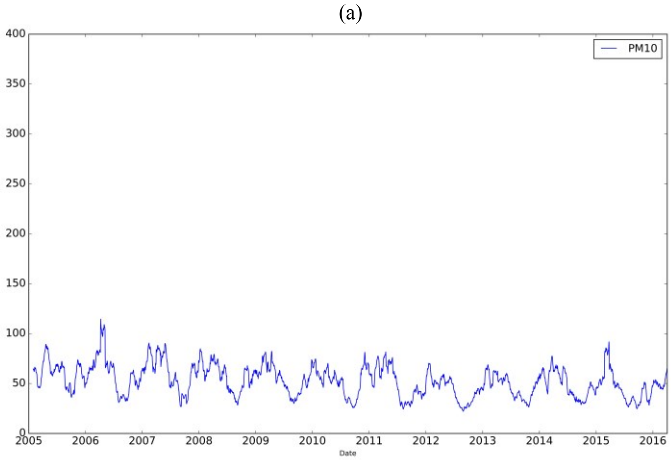
LSTM has an input gate, output gate, and forget gate. Sequence data created previously enters through the input gate of PM10 forecasting LSTM model. Data coming in through the input gate uses mini-batch to learn not all the data but a part of the data repeatedly.

The output gate outputs through ReLU the value of the next day forecast with data across 30 days. ReLU is normally used for image recognition. ReLU maintains the values greater than 1 as linear values and outputs the values that are less than or equal to 0 as 0. Hence, this paper uses ReLU to output data greater than 0 when learning.

RNN is calculated from the past concealed value, but data long past is difficult to calculate the gradient from. This phenomenon is known as *vanishing gradients,* and the forget gate among LSTM solves the vanishing gradients problem by only allowing a part of data to be used.



(a)



(b)

Fig. 1. (a): Raw data, (b): Moving Average data

sequence \ time	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$	$t+9$
1	21	15	17	18						
2		15	17	19	20					
3			17	19	25	24				
4				...	...	...	...			
5										
6										
7										

Input data | Predict data

Fig. 2. Sequence data example

The RMSE value is measured by changing the values of the parameter list in Table 1. Finally, the parameter used for the PM10 forecast LSTM model is shown in Table 1.

In order to optimize the model, RMSProp is used. RMSProp is normally used to optimize non-stationary time series [26]. Non-stationary time series means not knowing where the time series would flow towards (e.g. stock market). Through the random parameter assigned by the user, RMSProp updates itself and acts to control the learning rate. Therefore, RMSProp is used for optimized models.

The tool used in this paper is Keras [27], which is a deep learning tool of Python. In addition, the LSTM algorithm among Keras' RNN models is used in studying forecasts from the PM10 data collected.

VI. EVALUATION

Section 6 explains the performance evaluation algorithm and the formation of the training set, the test set, and the validation set. Performance comparisons between linear regression, RNN, and Adagrad model of LSTM are also conducted. The algorithm used in performance evaluation is Root Mean Square Error (RMSE) and Mean Square Error (MSE). The result of performance evaluation shows that LSTM model

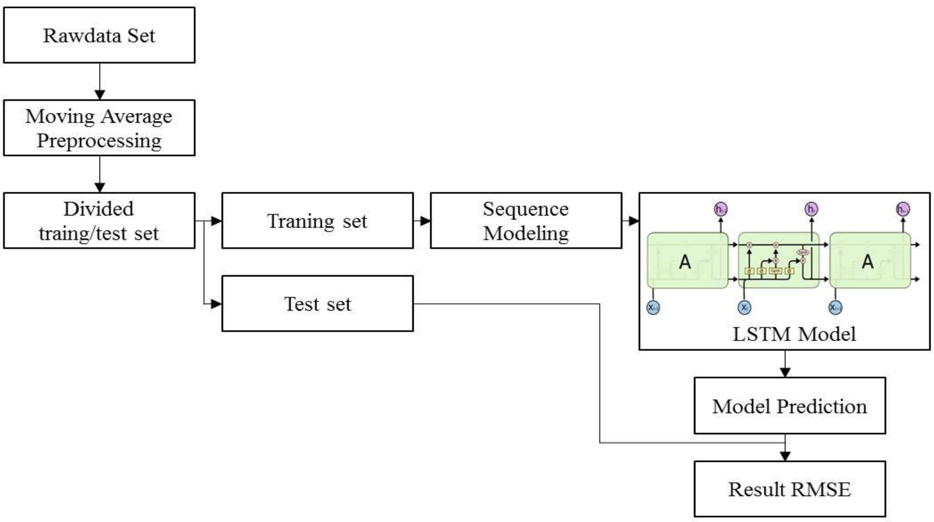


Fig. 3. Flow chart of the suggested PM10 forecast LSTM Model

Table 1. LSTM Parameter Settings

Index	Parameters					
	Length of Sequence	Hidden Node	Batch Size	Epoch	MSE	RMSE
1	30	500	7	20	1578.23	40.53
2	30	500	7	30	4.85	2.50
3	30	500	7	40	4.02	2.02
4	**30**	**500**	**7**	**50**	**3.53**	**1.86**
5	30	500	7	60	3.64	1.91

Index	Parameters					
	Length of Sequence	Hidden Node	Batch Size	Epoch	MSE	RMSE
1	30	100	30	50	20.23	10.12
2	30	300	30	50	6.14	2.48
3	30	300	7	50	4.71	2.17
4	30	500	30	50	4.02	2.18
5	**30**	**500**	**7**	**50**	**3.53**	**1.86**

using RMSProp showed 500% better performance than linear regression and 100% better than the RNN model.

The total number of units of data used in the PM10 forecast is 4,079, which is the total number of units of data--4,108--subtracted by 30 units of data after MA preprocessing. The training set and test set are configured by using the repeated sequence data. The training set used 3261 units of data, whereas the validation set uses 98 units of data, which is 3% of the training set. 818 units of data are configured and used as the test set.

In order to compare the performance of the LSTM model with that of other algorithms, the linear regression model, RNN model and Adagrad model of LSTM are used for comparison. Three models used the training set to learn in sequential order, and when learning with the validation set, the training set is also learned repeatedly. Lastly, the test set is used to forecast on the model generated by the training set.

MSE and RMSE are used to evaluate. These algorithms are good methods to view the differences between the forecast data generated by a model and real-life data. For RMSE, if the number is closer to 0, it represents better performance.

$$RMSE \sqrt{\frac{\sum_{i=1}^{n}(X_{obs_i} - X_{model_i})^2}{n}}$$

In (4), X_{obj} is the real value, whereas X_{model} is the value forecast by using the PM10 LSTM forecast model in time i.

The models used for performance review are linear regression, RNN, and the LSTM (Adagrad [8], RMSProp) model. The Adagrad optimization function of LSTM is one of the most used optimization functions in non-stationary time series, together with the RMSProp optimization function. Therefore, it is used to compare and evaluate against the RMSProp function of the LSTM model. The test results are shown in Table 2.

Fig. 4 represents the changes in RMSE value and it is called LSTM loss. This refers to the error value shown when applying the model learned only with training set in LSTM learning to the test set. By repeatedly learning the loss, the RMSE value of the LSTM model is minimized.

This paper uses the LSTM model to forecast PM10. The model was evaluated by comparing it with other research models like ANN and linear regression.

Table 2. Model Evaluation

	MSE	**RMSE**
Linear Regression	105.54	10.27
RNN	3.71	1.96
LSTM (RMSProp)	**3.53**	**1.86**
LSTM (Adagrad)	3.91	1.98

VII. CONCLUSION

The research result using data presented in this paper showed that LSTM improved the performance of RMSE better than linear regression by 500%. Upon comparing by using the Adagrad-optimized function of the RNN and LSTM, the RMSProp optimized model showed better performance by 100%.

However, the current result is a suggested paper on univariate PM10 data using the LSTM model. Hence, it is expected that further research will be required on the LSTM forecast model with additional factors like climate and disease. Furthermore, it is expected to forecast fine particulate matter (PM2.5) by adding various factors.

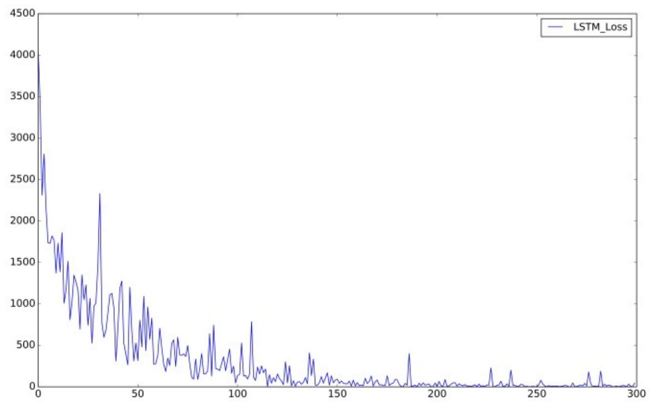


Fig. 4. LSTM loss

VIII. ACKNOWLEDGEMENT

This study was supported by Amway Korea, Ltd. (Seoul, Korea) The funding sources had no involvement in the collection, analysis, and interpretation of the data; the writing of this report; or the decision to submit this manuscript for publication and this study was supported by the ICT R&D program of MSIP/IITP, Korea.[R0601-16-1039, Building an Infrastructure of a Large Scale Data Center].

IX. REFERENCES

[1] S. Hochreiter, J. Schmidhuber, “Long short-term memory “, Neural Computation, vol.9, no.8, pp. 1735-1780, November 1997.

[2] http://www.airkorea.or.kr/index

[3] M. S. Souza, P. Coelho, A. da Silva, A. Pozza, “Using Ensembles of Artificial Neural Networks to Improve PM10 Forecasts”, Chemical Engineering Transactions, vol 43, pp. 2161-2166, 2015

[4] B. Oancea, S. C. Ciucu, “Time series forecasting using neural networks”, Proceedings of the CKS 2013 International Conference, pp. 1402-1408, January, 2014

[5] https://en.wikipedia.org/wiki/Recurrent_neural_network

[6] V. Nair, G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines”, International Conference on Machine Learning, , pp. 807-814, 2010

[7] http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

[8] J. Duchi, E. Hazan, Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”, Journal of Machine Learning Research, vol. 12, pp. 2121-2159, 2011

[9] T. Chai, R.R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature”, Geoscientific Model Development, vol. 7, pp. 1247-1250, June 2014

[10] B. K. Mutai, N. J. Muthama, J. Nganga, J. N. Ngaina, “Weather-dependence of fine particulate matter air quality over kenya”, The 11th Kenya Meteorological Society International conference, November 2013

[11] D. Dunea, S. Iordache, C. Ianache, “Relationship between airborne particulate matter and weather conditions in targoviste urban area during cold months”, Roumanian Journal of Chemistry, vol. 60, pp. 595-601, 2015

[12] MM. Suarez-Varela, A. Gallardo-Juan, L. Garcia-Marcos, N. Cimeno-Clemente, “The impact of atmospheric pollutants on the prevalence of atopic eczema in 6-7-year-old schoolchildren in Spain”, ISAAC Phase III. Iran J Allergy Asthma Immunol, vol. 12, pp 220-227, July 2013

[13] WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide, 2005

[14] Unitied States Environmental Protection Agency, NAAQS Table

[15] D. Vlachogiannis, A. Sfetsos, “Time series forecasting of hourly PM10 values: model intercomparison and the development of localized linear approaches”, WIT Transactions on Ecology and the Environment, vol. 86, pp. 85-94, 2006

[16] S. Thomas, B. Jacko, “Model for Forecasting Expressway Fine Particulate Matter and Carbon Monoxide Concentration: Application of Regression and Neural Network Models”, Journal of the Air&Waste Management Association, vol. 57, pp 480-488, April 2007

[17] M. Misity, Y. Misiti, J. M. Poggi, B. Portier, “Mixture of linear regression models for short term PM10 forecasting in Haute Normandie(France)”, Case Studies In Business, Industry, And Govemment Statistics, vol. 6, pp. 47-60, August 2013

[18] H. A. Hamid, A. S. Yahaya, N. A. Ramli, A. Z. U. Saufie, M. N. Yasin, “Short Term Prediction of PM10 Concentrations Using Seasonal Time Series Analysis”, MATEC Web of Conferences,vol. 47, 2016

[19] H. Weizhen, L. Zhengqiang, Z. Yuhuan, X. Hua, Z. Ying, L. Kaitao, L. Donghui, W. Peng, M. Yan, “Using support vector regression to predict PM10 and PM2.5”, In IOP Conference Series: Earth and Environmental Science, vol. 17, 2014

[20] Y. Zhao, Y. A. Hasan, “Machine learning algorithms for predicting roadside fine particulate matter concentration level in Hong Kong Central”, Computational Ecology and Software, vol. 3, pp. 61-73, 2013

[21] C. Lipton,C. Kale, C. Elkan, R. Wetzel, “Learning to diagnose with lstm recurrent neural networks”, International Conference on Learning Representations, November 2015

[22] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, “Long Short Term Memory Networks for Anomaly Detection in Time Series”, European Symposium on Artificial Neural Netwokrs, pp. 89-94, April 2015,

[23] A. Gers, D. Eck, J. Schmidhuber, “Applying LSTM to Time Series Predictable Through TimeWindow Approaches”, ICANN Proceedings of the International Conference on Artificial Neural Networks, August 2001, pp 669-676

[24] J. Chung, C. Gulcehre, K. H. Cho, Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”, Neural Information Processing Systems , December 2014

[25] http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[26] https://en.wikipedia.org/wiki/Stationary_process

[27] https://keras.io/