INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 5)
Available online at: www.ijariit.com

Prediction of Indoor $PM_{2.5}$ concentrations using Support Vector Regression

Satya Sundar Patra
spatra00@gmail.com
Indian Institutes of Technology, Chennai, Tamil Nadu

ABSTRACT

Studies have found that exposure to elevated levels of $PM_{2.5}$ concentrations, both in the long term and short term, has adverse health effects. The risk of exposure in the indoor microenvironment is higher compared to that of outdoor because people spend a significant amount of their time indoors. Therefore, to mitigate the health risks, it becomes essential to maintain indoor $PM_{2.5}$ levels. To do so, keeping an eye on the future expected concentrations is equally important as monitoring the current $PM_{2.5}$ levels. This paper talks about the development of one such model that enables us to predict the indoor $PM_{2.5}$ concentrations using Support Vector Regression. Upon development the model showed a very high value of R-square (0.98) and while testing it yielded a very low error (3.78%).

Keywords— *Indoor $PM_{2.5}$, $PM_{2.5}$ Prediction, Support Vector Regression (SVR) and Low-cost sensor*

1. INTRODUCTION

Exposure to fine particulate matter (particulate matters having aerodynamic diameter ≤ 2.5µm; $PM2_{2.5}$) has been reported to be associated with increased mortality (Pope et al. 2002; Franklin et al. 2007; Zanobetti et al. 2009; Shi et al. 2016). Epidemiological studies have illustrated that $PM2_{2.5}$ affects the respiratory system, the cardiovascular system, the nervous system, and also the immune system (Koike et al. 2008; Ranft et al. 2009; Van et al. 2010; Lim et al. 20110). Besides health effects, it also disturbs the economic, and social development considerably in the long term (Cao et al. 2018). A subtle change in $PM_{2.5}$ levels might also have significant effects. A study reported that for every 10 µg/m^3 change in PM levels, the admission rate of coronary heart disease increased by 1.89%, occurrence of heart attack increased by 2.25%, morbidity of congenital coronary heart disease increased by 1.85%, and risk of respiratory disease increased by 2.07% (Forastiere et al. 2008). Studies in European countries have estimated nearly 4,00,000 premature deaths occur due to $PM2_{2.5}$ exposure (Badyda et al. 2017). In India, the present PM pollution in Delhi registered 7350–16,200 premature deaths and nearly 6.0 million asthma attacks annually (Guttikunda et al. 2013).

These numbers corresponding to the ill-effects of $PM2_{2.5}$ pollution is disturbing. Therefore, it becomes crucial to regulate the $PM_{2.5}$ levels in the surroundings we live in. A study reported by EPA suggested that an average American spends 87% of their life indoors. The number of deaths due to indoor air pollution published by the World Health Organization (WHO) touches nearly 3.8 million annually (WHO, 2019). Thus, indoor microenvironments become one of the most critical sites for PM monitoring.

In addition to monitoring the current $PM_{.2.5}$ levels, it is equally important to keep a close eye on future expected concentrations. This is because a forecast close to the observed value would help people to access the forthcoming variation in the levels and thereby prepare accordingly. However, not many studies have been reported in this regard. To fill in this gap, the current study aims to propose a methodology to predict future indoor $PM2_{2.5}$ concentrations using Support Vector Regression. This would benefit people with knowledge of future expected exposure patterns and help considerably to mitigate the risks associated with exposure.

2. MATERIALS AND METHODS

2.1 $PM_{2.5}$ measuring sensor

A low-cost portable PM measuring sensor, Sensiron SPS30, is used in this study to measure the $PM2_{2.5}$ concentrations. SPS30 counts the number of particles by employing a light scattering technique. The intensity of the light scattered after it incidents on an aerosol particle is proportional to the size of the particle. In this way, the sensor counts the number of particles in different size bins. The number concentration is then converted into mass concentration by assuming an appropriate density of the aerosol particles (Hinds et al. 1999). SPS30 was particularly selected because it comes pre-calibrated using potassium chloride salt particles and the TSI DustTrekTM DRX Aerosol Monitor 8533 as a reference (Sensiron, 2018). The technical specifications of the sensor are listed in Table 1.

Table 1: Sensiron SPS30 specifications

Power Requirements	4.5-5.5 V, <60mA
Range of the instrument	1-1000 μg/m^3
The lower limit of the sensor	0.3 μm
Minimum scanning interval	1 sec
Dimensions	40.6 x 40.6 x 12.2 mm

However, SPS30 does not come with an in-build data storage unit. Hence to log the data, it is interfaced with a Raspberry Pi4 using I2C communication. Raspberry Pi4 is a single-board computer with a Quad-core 64-bit ARM-Cortex A72 processor running at a bandwidth of 1.5GHz (Raspberry Pi, 2019). It comes with an in-built ram of 4GB and supports external ROM using a micro-SD slot. Python code is written to read the output of the I2C communication and store it into a text file. The setup of the SPS30 and the raspberry pi is shown in figure 1.

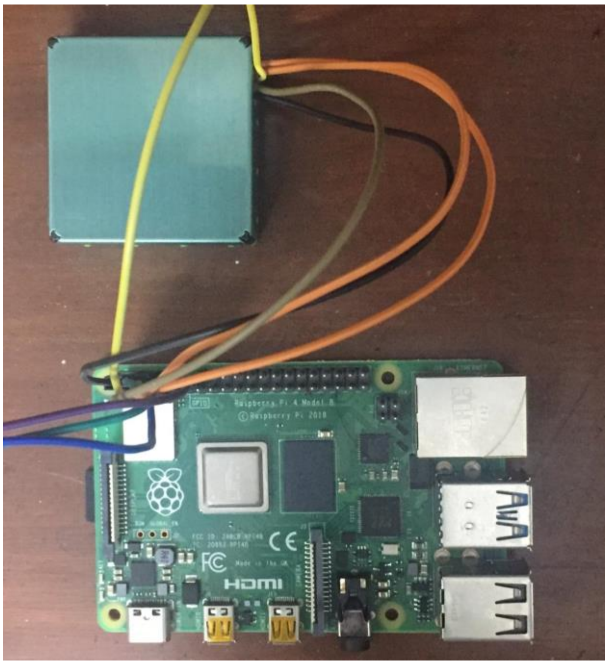

Fig. 1: SPS30 connected with Raspberry Pi4

This setup was placed in a single-occupancy indoor microenvironment in Chennai, India to measure $PM_{2.5}$ for a duration of 24 hours. The sampling interval of 5 minutes was chosen for the study.

2.2 Prediction Model

The prediction function proposed is expressed in Equation 1.

Where $(PM_{2.5})_T$ is the current $PM_{2.5}$ concentration

$$(PM_{2.5})_{T+1} = f((PM_{2.5})_T, (PM_{2.5})_{T-1}, (PM_{2.5})_{T-2}, (PM_{2.5})_{T-3}, (PM_{2.5})_{T-4}) \qquad \text{(Equation 1)}$$

$(PM_{2.5})_{T+1}$ is the next interval $PM_{2.5}$ concentration which is to be predicted
$(PM_{2.5})_{T-1}$, $(PM_{2.5})_{T-2}$, $(PM_{2.5})_{T-3}$, $(PM_{2.5})_{T-4}$ are the previous four interval's $PM_{2.5}$ concentrations

Here, the future to-be-predicted $PM_{2.5}$ value is expressed as a function of current and previous four interval's $PM_{2.5}$ concentration. The function is however not of a known kind. Hence, a non-parametric regression, Support Vector Regression (SVR) (Smola et al. 2004) is used. SVR finds a function *f(x)* which is at-most ϵ deviation from observed values in the training data set, at the same time tries to make the function as flat as possible. To do so, SVR uses kernel functions. In this study, a linear kernel is used. The function for linear SVR takes the form,

$$f(x) = w\,.\,x + b \qquad \text{where } w, b \subset R \qquad \text{(Equation 2)}$$

The flatness of the function means a smaller value of *w*. This function *f(x)* for linear SVR is determined by solving the optimization problem given in Equation 3.

$$\text{minimize} \quad \frac{1}{2}||w||^2 + C(\sum \zeta + \zeta^*)$$

$$\text{subject to} \begin{cases} y_i - w.x_i - b \leq \epsilon + \zeta \\ w. + b - y_i \leq \epsilon + \zeta^* \\ \zeta, \zeta^* \geq 0 \end{cases} \qquad \text{(Equation 3)}$$

Where ζ, ζ^* are the slack variables, introduced to avoid infeasible optimization. The constant *C* determines the tradeoff between the flatness and ϵ. The solution of the optimization problem determines the function which is used for prediction. In the current work, the SVR model is developed in the regression learner app in Matlab.

3. RESULTS

3.1 Training

The collected data was divided into two sets, training, and testing. The training set comprised 75% of the data and the rest formed the testing set. Data from the training set was used to build the SVR model. To avoid over-fitting fivefold cross-validation was enabled while training. The Response plot of training is shown in figure 2.

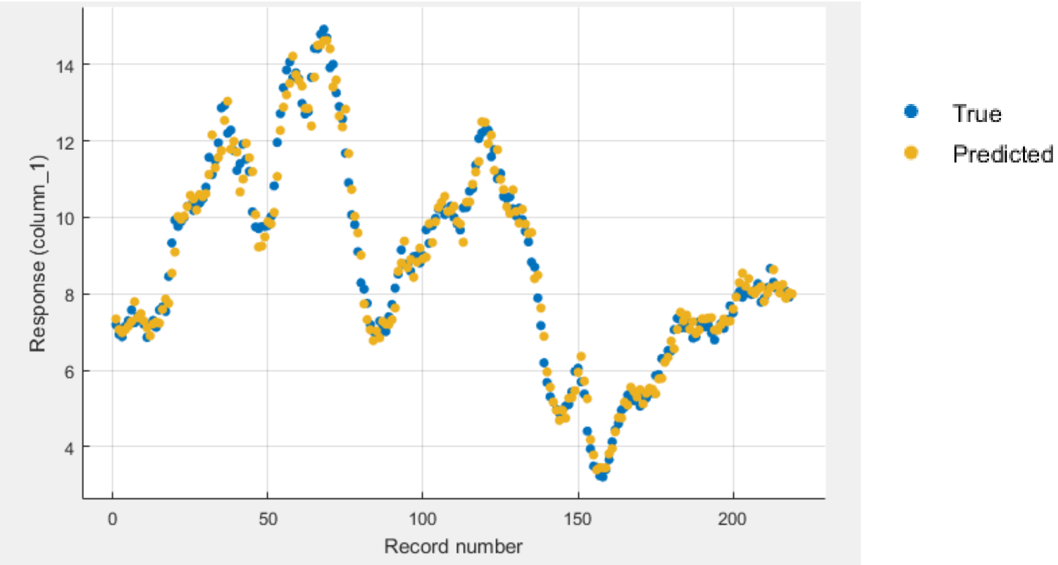

Fig. 2: Response plot of training

The adjusted R-squared value of the model is 0.98, which indicates that about 98% of the variability in the data can be explained by this model.

3.2 Testing

The developed SVR model was tested for the rest of 25% of the data. The comparison of the actual value and the predicted value is represented in figure 3.

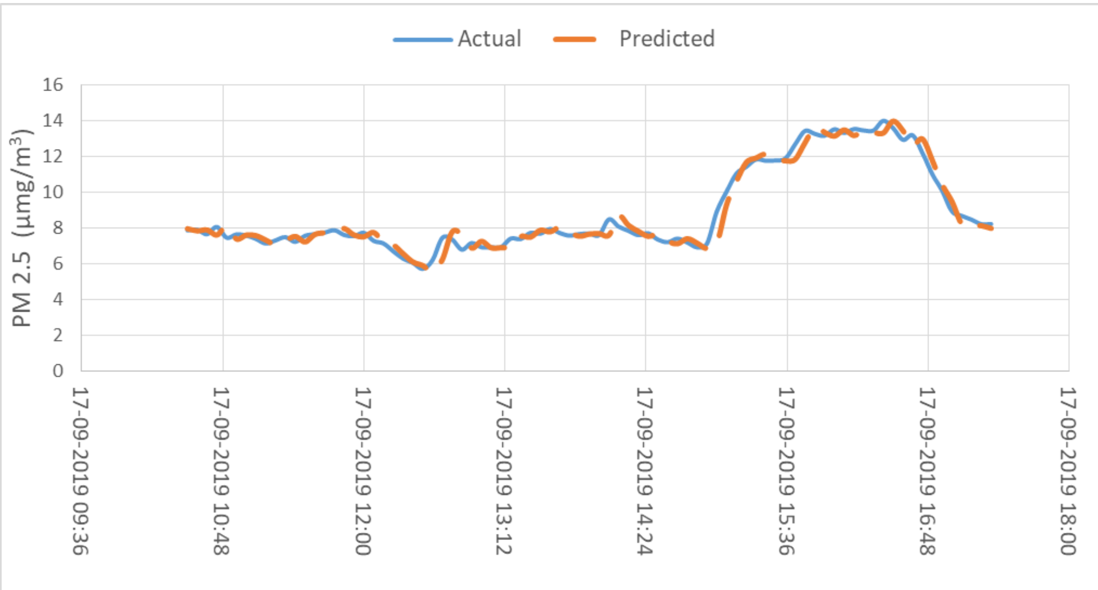

Fig. 3: Comparison of predicted vs. actual $PM_{2.5}$ concentration for the test data

Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are used to quantify the performance of the model. Mathematically,

$$MAPE = \frac{|(y_i - y^*_i)|}{y_i}$$

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

where Y_i is the observed value and Y^*_i is the predicted value.

The MAPE for the testing data is 3.78% and the RMSE is 0.44µg/m^3. The low values of MAPE and RMSE while testing indicates a good functionality of the prediction model.

4. CONCLUSION

This paper proposes a model to predict indoor $PM_{2.5}$ concentrations. The proposed model is developed using Support Vector Regression (SVR) with current and past four observed $PM_{2.5}$ values as input. A very high value of adjusted R-square (0.98) is observed while training the model. Upon testing, the forecasted values were found to be close to the observed values. The Mean Absolute Percentage Value (MAPE) observed during testing was 3.78% and Root Mean Squared Error was 0.44 µg/m^3. A high value of adjusted R-square and low values of MAPE and RMSE suggests that the proposed model using the selected inputs and SVR articulates an outstanding performance for $PM_{2.5}$ predictions.

5. REFERENCES

[1] Pope III, C. Arden. 2002. "Lung Cancer, Cardiopulmonary Mortality, And Long-Term Exposure To Fine Particulate Air Pollution". JAMA 287 (9): 1132. doi:10.1001/jama.287.9.1132.

[2] Franklin, Meredith, Ariana Zeka, and Joel Schwartz. 2006. "Association Between PM2.5 And All-Cause And Specific-Cause Mortality In 27 US Communities". Journal Of Exposure Science & Environmental Epidemiology 17 (3): 279-287. DOI:10.1038/sj.jes.7500530.

[3] Zanobetti, Antonella, and Joel Schwartz. 2009. "The Effect Of Fine And Coarse Particulate Air Pollution On Mortality: A National Analysis". Environmental Health Perspectives 117 (6): 898-903. DOI:10.1289/ehp.0800108.

[4] Shi, Liuhua, Antonella Zanobetti, Itai Kloog, Brent A. Coull, Petros Koutrakis, Steven J. Melly, and Joel D. Schwartz. 2016. "Low-Concentration PM 2.5 And Mortality: Estimating Acute And Chronic Effects In A Population-Based Study". Environmental Health Perspectives 124 (1): 46-52. DOI:10.1289/ehp.1409111.

[5] Koike, Eiko, Hirohisa Takano, Ken-ichiro Inoue, Rie Yanagisawa, and Takahiro Kobayashi. 2008. "Carbon Black Nanoparticles Promote The Maturation And Function Of Mouse Bone Marrow-Derived Dendritic Cells". Chemosphere 73 (3): 371-376. DOI:10.1016/j.chemosphere.2008.05.054.

[6] Ranft, Ulrich, Tamara Schikowski, Dorothee Sugiri, Jean Krutmann, and Ursula Krämer. 2009. "Long-Term Exposure To Traffic-Related Particulate Matter Impairs Cognitive Function In The Elderly". Environmental Research 109 (8): 1004-1011. DOI:10.1016/j.envres.2009.08.003.

[7] Van Donkelaar, Aaron, Randall V. Martin, Michael Brauer, Ralph Kahn, Robert Levy, Carolyn Verduzco, and Paul J. Villeneuve. 2010. "Global Estimates Of Ambient Fine Particulate Matter Concentrations From Satellite-Based Aerosol Optical Depth: Development And Application". Environmental Health Perspectives 118 (6): 847-855. DOI:10.1289/ehp.0901623.

[8] Lim, J.-M., J.-H. Jeong, J.-H. Lee, J.-H. Moon, Y.-S. Chung, and K.-H. Kim. 2010. "The Analysis Of PM2.5 And Associated Elements And Their Indoor/Outdoor Pollution Status In An Urban Area". Indoor Air 21 (2): 145-155. DOI:10.1111/j.1600-0668.2010.00691.x.

[9] Cao, Qilong, Guoqiang Rui, and Ying Liang. 2018. "Study On PM2.5 Pollution And The Mortality Due To Lung Cancer In China Based On Geographic Weighted Regression Model". BMC Public Health 18 (1). DOI:10.1186/s12889-018-5844-4.

[10] Forastiere, Francesco, Massimo Stafoggia, Giovanna Berti, Luigi Bisanti, Achille Cernigliaro, Monica Chiusolo, and Sandra Mallone. 2008. "Particulate Matter And Daily Mortality". Epidemiology 19 (4): 571-580. doi:10.1097/ede.0b013e3181761f8a.

[11] Badyda AJ, Grellier J and Dąbrowiecki P. 2017. "Ambient PM2. 5 exposure and mortality due to lung cancer and cardiopulmonary diseases in Polish cities. Adv Exp Med Biol.; 944:9–17. https://doi.org/10.1007/978-3-319-44488-8_55

[12] Guttikunda, Sarath K., and Ramani V. Kopakka. 2013. "Source Emissions And Health Impacts Of Urban Air Pollution In Hyderabad, India". Air Quality, Atmosphere & Health 7 (2): 195-207. DOI:10.1007/s11869-013-0221-z.

[13] "WHO". 2019. World Health Organization. https://www.who.int/airpollution/en/.

[14] Hinds, W. C. (1999). Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles, 2nd Ed., New York: Wiley-Interscience.

[15] "Sensirion". 2018. Sensirion.Com. https://www.sensirion.com/en/environmental-sensors/particulate-matter-sensors-pm25/.

[16] Raspberry Pi. 2019. https://www.raspberrypi.org/products/raspberry-pi-4-model-b/specifications/.

[17] Smola, Alex J., and Bernhard Schölkopf. 2004. "A Tutorial On Support Vector Regression". Statistics And Computing 14 (3): 199-222. doi:10.1023/b:stco.0000035301.49549.88.

