International Journal Of Artificial Intelegence Research
Vol 6, No 1, June 2022
ISSN: 2579-7298

Implementation of Tensor Flow in Air Quality Monitoring Based on Artificial Intelligence

Untung Rahardja [a,1], Qurotul Aini [a,2], Danny Manongga [b,3], Irwan Sembiring [c,4], Iqbal Desam Girinzio [a,5]

[a] University of Raharja, Tangerang, [b] Satya Wacana Christian University, Salatiga , [c] Satya Wacana Christian University, Salatiga
[1] Untung@raharja.info, [2]aini@raharja.info, [3] danny.manongga@uksw.edu, [4] irwan@uksw.edu, [5] iqbal.desam@raharja.info
* corresponding author

ARTICLE INFO

Article history:
Received 20 Feb 2022
Revised 05 Apr 2022
Accepted 18 June 2022

Keywords:
TensorFlow
Artificial Intelligence
Air Quality
Monitoring
Air Pollution

ABSTRACT

Chemicals that cannot be controlled today can pollute resources and the environment. Common sources of pollutants are due to public transportation, cigarette smoke, volcanic activity that emits volcanic ash, factory smoke, forest fires, biogas, or carbon dioxide. The purpose of this paper is to monitor air quality, detect air and anticipate pollution levels. With the specified algorithms, three algorithms will be used to create a good and accurate model where four different gasses are predicted: carbon dioxide, sulfur dioxide, and nitrogen dioxide, in this paper, there are four algorithms used for the Air Qualification Index which are Support Vector Regression, Linear Regression, and Ensemble Gradient Boosted Decision Tree. This research also includes quantitative research which is hypothesized to be evaluated against Root Mean Squared Error, Mean Squared Error, and Mean Absolute error, depending on the performance of the measurements made by artificial intelligence, and the lower error value is selected. Based on the algorithm to be predicted in this air quality monitoring, there are 5 air pollutants like Carbon dioxide, Sulfur dioxide, and Nitrogen dioxide, and the sensors to be used are two sensors like PM2.5 and PM10 that can be predicted.

I. Introduction

Air is an important factor in life. Air quality that has been contaminated with chemicals can threaten the survival of all creatures ranging from humans, animals to plants [1]. Various dangerous air containing emissions such as public vehicle emissions, industrial emissions such as the coal industry, and emissions from active volcanoes [2]. There are many illnesses brought on by air pollution, including heart attacks, lung conditions, chronic bronchitis, and respiratory illnesses, but it can also lead to immediate issues like sneezing and shortness of breath [3]. Studies say that people living in urban areas are at risk of developing lung disease [4].

There are numerous direct health implications of air pollution. Industry, traffic, energy power plants, and the use of solid fuels in residential buildings all have a substantial impact on air pollution [5]. Increasing pollution poses a grave threat to the economy and the general public's quality of life. The most used method of measuring pollution is the Air Quality Index, which combines five significant contaminants.

Particulate matter, carbon monoxide, sulfur dioxide, nitrogen dioxide, and ground-level ozone. As a result, air pollution is more likely to occur presently. This necessitates immediate action and preventive measures in order to not only stop the growth in air pollution but also to Therefore, It's crucial to have technologies that can forecast future pollution levels in addition to monitoring current pollution. By locating pollution sources, AI can track and forecast changes in air pollution levels. Think about whether greater industrial production is correlated with better air quality, or whether constraints on transportation are a factor. Air pollution monitoring stations, which are big and expensive to establish and operate, are usually used to measure air quality.

However, the data from these sites is very accurate in terms of the quality of the air. Other techniques are reported to be more affordable and to cover a larger region. Industry and academia are both interested in this relatively new technology called the Internet of Things. To solve the weaknesses of present air quality measurement systems in recognizing and predicting chemicals that become very close to the environment and to reduce the total cost of air pollution, the research offers a novel technique. In exchange, reporting and predicting tools driven by AI are dependable, affordable, straightforward to use, and quick to implement. The level of skill that AI needs to produce these predictions, in particular, has no hard upper limit. The field of artificial intelligence expands as new findings are made in it. Linear regression, support vector regression, and gradient boosted decision tree for prediction are three artificial intelligence learning techniques that are defined.

This project will also use a library from python, TensorFlow, Tensorflow is a tensor and computational graph that crosses nodes to edges. So, Data for Machine Learning generally requires computation in some form and therefore, we often see data representation done numerically. To define a tensor: a container that can hold data in N dimensions. In short, a tensor is a mathematical object used to describe physical properties such as scalars and vectors. Basically, tensors are just a generalization of scalars and vectors; scalars are zero-rank tensors, while vectors are first-rank tensors..

II. Methods

This research employs artificial intelligence approaches to evaluate and estimate sensor concentrations and ecological implications in order to examine the probable concentration of gaseous pollutants in the environment and alert locals and appropriate authorities [11]. The system includes a monitoring and warning mechanism that operates in real-time [12]. Future forecasters apply these calculated parameters to send emergency alerts using artificial intelligence learning algorithms, Support vector regression, specifically linear regression, and gradient boosted tree decision.

Here are the algorithms used to create this artificial intelligence:

A. Support Vector Regression

Based on the support vector theory developed by Vapnik, support vector regression is a supervised learning procedure [13]. To lessen uncertainty, it is necessary to locate hyperplanes and decrease the difference between expected and actual values. For the computation below, the minimal value of "S" equals the maximum margin theorem, where the formula specifies the percentage of total empirical error.

B. Linear Regression

Predictive analytics and the widely used linear regression model are the most fundamental methods [14]. The linear technique and the relationship between its two variables are relevant. MLR, often referred to as simple multiple regression, is a quantitative method for forecasting the results of a dependent variable by using a number of important independent factors. The purpose of multiple regression is to describe the linear relationship between the causative (independent) and response (dependent) elements (MLR).

C. Gradient Boosted Decision Tree

GBDT develops an ensemble learner using the boosting technique. To create a powerful learner, decision trees are connected sequentially [15]. The GBDT decision trees do not fit the complete dataset. The objective is to reduce the errors in the previous tree. Every tree thus conforms to the residuals of the previous tree. As a result, the model eventually becomes more accurate and robust overall.

To connect intelligent objects to the Internet, the OSI layer protocol's capabilities are used as the Web of Things' fundamental purpose. In this work, we propose mobile nodes for such levels of air performance sensors that monitor levels of air pollutants [16]. All Air Sensors interface is a tiny sensor connection system which connects to the web or, by extension, a worldwide network of linked objects. Data from the Internet of Things is gathered using an artificial intelligence platform, and all of the sensor data is saved and sent to the cloud.

1) Data collection on air quality:

The Internet of Things device was built utilizing various components, including gas sensors, embedded systems, Raspberry gadgets, clean air sensor systems, and $PM_{2.5}$ particle sensors. Although

$PM_{2.5}$ particle sensors locate nanomaterials with such a size smaller than 2.5 microns, air pollutants monitors compute the Index of Air Quality (AQI) [17]. Gas sensors are used to monitor the amounts of nitrogen gas (NO_2), carbon dioxide (CO_2), and sulfur dioxide (SO_2). The required data is gathered by a sensor, which is connected to a microcontroller for reading. This information is gathered and then sent to the cloud via MCU nodes. The acquired data can be used to forecast the upcoming few hours using the Index of Air Quality (AQI) as a test dataset.

2) Pre-processing:

Multiple sources of data may contain redundant data, missing figures, and contradicting formations [18]. To obtain the proper prediction performance, the dataset must be cleaned; missing values must be filled in, deleted, or handled using another method that uses an average value. To avoid skewed outcomes, redundant data should also be deleted or discarded. Some data sets might contain abnormalities or extreme values that should be eliminated to produce a fair estimate. Only when all of this data preprocessing is completed can classification algorithms and other sorts of data mining be effective. For data pre-processing and to minimize some redundant information, classification methods are also used after logistic regression.

3) Air Quality Measurement

Important variables in the suggested methodology include

a. *Carbon Dioxide (CO_2)*: Carbon dioxide is an inflammable, flavorless, and colorless gas [19]. It also examines the category for asphyxiating gasses that block tissues' access to oxygen. CO_2 is a very important emission for the planet, as it is one of the main components of the photosynthesis process, changing chemical carbon dioxide energy to solar carbon. A significant contributor to the growing CO_2 levels is the increased usage of fossil fuels. Plant development is hastened as a result. Because unwanted vegetation spreads quickly, pesticide use rises..
b. *Sulfur Dioxide (SO_2)*: Sulfur oxides are a wide class of compounds that include poisonous, colorless glasses (SOx). When fossil fuels (coal, oil, diesel) or other substances containing sulfur are burned, these glasses, particularly SO_2, are produced. Power plants, smelters, and cars are a few examples of sources. Sulfur dioxide emissions from diesel machinery and automobiles have historically been a significant source, but recent federal rules to lower the sulfur content of diesel fuel have dramatically reduced emissions from this industry. Another unavoidable result of volcanic activity is sulfur dioxide. Sulfur dioxide is a colorless gas that can be detected by taste and smell. The main culprits are fossil fuels.
c. *Nitrogen Dioxide (NO_2):* The collection of related gasses known as nitrogen oxides, or NOx, includes nitrogen dioxide, A gaseous air pollutant consisting of both oxygen and nitrogen frequently referred to as NO_2. NO_2 is produced when fossil coal, oil, fuel, gasoline, and biodiesel are burned at high temperatures [20]. The production of ozone and particle pollution are both influenced by the presence of NO_2 and other oxides of nitrogen in the surrounding air. This represents one of the six major air pollutants whose levels are restricted by federal air quality rules. Burning fossil fuels indoors, such as wood and natural gas, can also produce NO_2. Brownish-colored nitrogen dioxide is a gas that can be easily identified by its odor. It comes about as a result of burning fossil fuels. No chemical process is typically employed to release NO_2 into the atmosphere. In excessive amounts, NO_2 can harm the lungs. It, like SO_2, mitigates acid rain..

III. Result and Discussion

The dataset generated by must be used to test the presented models.

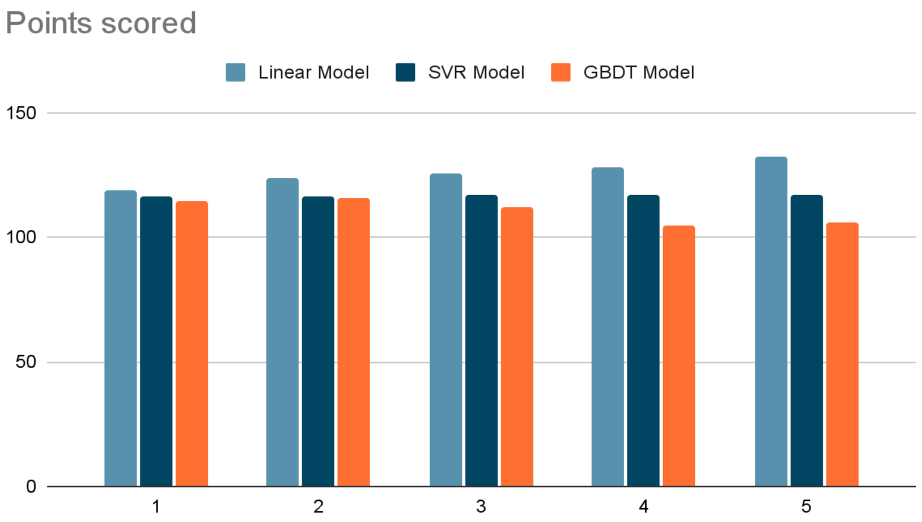

Fig. 1. AI model predictions for the Air Quality Index values

Shows how IoT can be utilized to predict the overall Air Quality Index for the coming hours. Burning fossil fuels indoors, such as wood and natural gas, can also produce NO2. Table 1 below shows that the predicted values are provided.

Table 1. Values of the AQI that models predict

Hour	Linear Model	SVR Model	GBDT Model
1	117.991	115.231	113.672
2	122.672	114.613	112.904
3	124.835	115.815	110.982
4	127.104	114.857	103.837
5	131.456	115.975	104.908

The proficiency statistics used in this study to assess how well the various models perform are RMSE, MAE, and MSE. A popular strategy for measuring model error in forecasting quantitative data is to separate the vector of expected outcomes from the vector of true values. Model performance values are presented in Table 2:

Table 2. Values for artificial intelligence model performance metrics.

Model	RMSE	MAE	MSE
Linear Regression	17.785	16.985	15.924
Support Vector Regression	08.455	10.977	7.914
Gradient Boosted Decision Tree	03.147	6.204	2.1405

This Journal uses a range of artificial intelligence approaches, including methods such as the support Vector Regression, Linear Regression and Gradient Boosted Decision Tree, to predict the Air quality monitoring. Examining performance measurement information for all models will allow one to choose Gradient Boosted Decision Tree with the smallest error value.

Validation datasets of different air pollution concentrations collected by monitoring systems were utilized to more easily evaluate the predicted accuracy of linear regression, support vector regression, and gradient boosted decision tree. Three metrics are also used to assess each model's performance: RMSE, MAE, and MSE. Table 3 compares the three methods used to predict the concentration of each pollutant and demonstrates that the gradient boosted decision tree works efficiently. in other words, has the best prediction accuracy and the least overall prediction error

Table 3. Error comparison between each algorithm.

Type	AI Model	MSE	RMSE	MAE
CO_2	Linear Regression	0.0019	0.0425	0.0917
	Support Vector Regression	0.0028	0.4565	0.0249
	Gradient Boosted Decision Tree	0.0004	0.2788	0.0169
SO_2	Linear Regression	0.4959	0.6989	0.4909
	Support Vector Regression	0.4586	0.7535	0.5925
	Gradient Boosted Decision Tree	0.0798	0.3115	0.3195
NO_2	Linear Regression	13.9016	5.9024	3.9018
	Support Vector Regression	10.9236	4.7824	3.0918
	Gradient Boosted Decision Tree	2.8926	1.5615	1.2377
$PM_{2.5}$	Linear Regression	24.9017	5.9166	3.7308
	Support Vector Regression	0.3425	0.6146	0.2917
	Gradient Boosted Decision Tree	0.0064	0.0915	0.0598
PM_{10}	Linear Regression	36.715	7.0244	3.9165
	Support Vector Regression	0.3427	0.6237	0.3169
	Gradient Boosted Decision Tree	0.0065	0.0727	0.0535

The metrics are used to assess the predictive model's efficacy and to look for any potential correlations between anticipated and actual values. Each of these metrics represents expected and actual values for the entire region, with the goal of forecasting values to solve the problem of air pollution. The metrics applied in this study are listed below.

1. Mean Squared Error (MSE):
   One of the simplest fundamental and simple measures used in regression is this one. It is defined as the mean squared error of the predictions produced, which is the sum of the squared differences between the true value and the predicted value. According to the formula One of the simplest and most fundamental measures used in regression is this one. It is referred to as the mean squared error of the predictions made or the sum of the squared deviations between the actual and expected values.
2. Root Mean Square Error (RMSE):
   Root mean square error is the process of calculating the square root of the mean square of the variations between the expected value and the actual value.
3. Mean Absolute Error (MAE):
   collects data of a divergence between two succeeding time series regardless the direction. This demonstrates that the strength of the mean and absolute deviations between predictions and data obtained is equal. Formula[20] displays the MAE formula.

IV. Conclusion

This paper describes an air production quality and forecasting model that makes use of forecasting techniques to offer a legitimate and suitable approach to the complexity of air quality. AI has the ability to accelerate international efforts to protect the climate and conserve resources by discovering energy sources that cut emissions. The forecasting component of the suggested system consists of a calculation method that resembles a sophisticated calculator. To determine how much pollution is produced and how it is transported across a specific location at a specific time, they use a number of mathematical calculations. The use of many sensors enhances the precision of pollution management, lowers the cost of observatories, and improves the efficiency and thoroughness of the monitoring of data about observation regions. In this study, predictions are made using a variety of models, including a gradient boosted decision tree, a vector regression support model, and a linear regression model. computational techniques, which resemble highly sophisticated calculators in their operation. They use a number of mathematical models to predict how much pollution will be released and how it will spread across a specific area at a specific time. Multiple sensors are used to ensure accurate pollution management, lower observatory costs, and data management over the effectiveness and size of the monitored zone. To predict the index of air quality values during the following five hours, a number of techniques are used, including linear regression, the Vectors Regression Support model, and gradient-boosted decision trees. An investigation of the performance criteria shows that the Gradient

Boosted Decision Tree does have the lower failure rate among all the models. Table 1 compares a number of models to the proposed system and displays projected values for the proposed model's anticipated values using AI techniques, clearly indicating the enhanced performance. Future studies may compare the timing and predicting capacities of various components to analyze the suggested IoT-equipped AI system. Additionally, it may be tested in a variety of circumstances and improved by dynamically adjusting the context in accordance with system operator considerations. More nodes with the ability to perform prediction algorithms could boost speed and dependability.

References

[1] S. R. Kumar, T. Arumugam, C. Anandakumar, S. Balakrishnan, and D. Rajavel, "Use of plant species in controlling environmental pollution," *Bull. Environ. Pharmacol. Life Sci.*, vol. 2, no. 2, p. 52, 2013.

[2] K. Ravindra, R. Sokhi, and R. Van Grieken, "Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation," *Atmos. Environ.*, vol. 42, no. 13, pp. 2895–2921, 2008.

[3] M. Roca, A. Verduri, L. Corbetta, E. Clini, L. M. Fabbri, and B. Beghé, "Mechanisms of acute exacerbation of respiratory symptoms in chronic obstructive pulmonary disease," *Eur. J. Clin. Invest.*, vol. 43, no. 5, pp. 510–521, 2013.

[4] A. S. Gershon, L. Warner, P. Cascagnette, J. C. Victor, and T. To, "Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study," *Lancet*, vol. 378, no. 9795, pp. 991–996, 2011.

[5] S. N. Sinha and P. K. Nag, "Air pollution from solid fuels," *Encycl. Environ. Heal.*, vol. 118, pp. 46–52, 2011.

[6] C. Huntingford, E. S. Jeffers, M. B. Bonsall, H. M. Christensen, T. Lees, and H. Yang, "Machine learning and artificial intelligence to aid climate change research and preparedness," *Environ. Res. Lett.*, vol. 14, no. 12, p. 124007, 2019.

[7] K. K. Khedo, R. Perseedoss, and A. Mungur, "A wireless sensor network air pollution monitoring system," *arXiv Prepr. arXiv1005.1737*, 2010.

[8] D. Müller, D. Klingelhöfer, S. Uibel, and D. A. Groneberg, "Car indoor air pollution-analysis of potential sources," *J. Occup. Med. Toxicol.*, vol. 6, no. 1, pp. 1–7, 2011.

[9] H. M. Taleb, "Using passive cooling strategies to improve thermal performance and reduce energy consumption of residential buildings in UAE buildings," *Front. Archit. Res.*, vol. 3, no. 2, pp. 154–165, 2014.

[10] S. Al-Janabi, M. Mohammad, and A. Al-Sultan, "A new method for prediction of air pollution based on intelligent computation," *Soft Comput.*, vol. 24, no. 1, pp. 661–680, 2020.

[11] N. Lutfiani, F. P. Oganda, C. Lukita, Q. Aini, and U. Rahardja, "Desain dan Metodologi Teknologi Blockchain Untuk Monitoring Manajemen Rantai Pasokan Makanan yang Terdesentralisasi," *InfoTekJar J. Nas. Inform. Dan Teknol. Jar.*, vol. 5, no. 1, pp. 18–25, 2020.

[12] U. Rahardja, Q. Aini, and A. Khoirunisa, "Monitoring Kinerja User Akuntan Menggunakan Dashboard Pada Web Based Accounting Online di Perguruan Tinggi," *Sains dan Teknol. Inf.*, vol. 4, no. 2, pp. 58–62, 2019.

[13] Q. Aini, N. Lutfiani, H. Kusumah, and M. S. Zahran, "Deteksi dan Pengenalan Objek Dengan Model Machine Learning: Model Yolo," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 6, no. 2, pp. 44–51, 2021.

[14] D. N. Sari and Q. Aini, "Geographically weighted bivariate zero inflated generalized Poisson

regression model and its application," *Heliyon*, vol. 7, no. 7, p. e07491, 2021.

[15] A. S. Bein, Y. I. Graha, and A. P. Pangestu, "Pandawan Website Design Based Content Management System As Media E-commerce Transaction," *Aptisi Trans. Technopreneursh.*, vol. 2, no. 1, pp. 87–97, 2020.

[16] A. Philip, C. S. Putri, and P. M. Arifanggi, "Traffic Light Timer Control Using Raspberry Pi," *Aptisi Trans. Technopreneursh.*, vol. 1, no. 2, pp. 134–143, 2019.

[17] F. Agustin, Q. Aini, A. Khoirunisa, and E. A. Nabila, "Utilization of Blockchain Technology for Management E-Certificate Open Journal System," *Aptisi Trans. Manag.*, vol. 4, no. 2, pp. 133–138, 2020.

[18] N. Yulandha, J. I. Saputro, and N. K. Nissa, "Design Information System Accounting Sales Website-Based (Case Study: PT Arbunco Wira Pandega)," *APTISI Trans. Manag.*, vol. 4, no. 2, pp. 157–167, 2020.

[19] A. S. Rafika, M. Hardini, A. Y. Ardianto, and D. Supriyanti, "Face Recognition based Artificial Intelligence With AttendX Technology for Student Attendance," in 2022 International Conferenc*e on Science and Technology (ICOSTECH)*, 2022, pp. 1–7.

[20] A. Faturahman, N. P. L. Santoso, W. Y. Prihastiwi, and B. A. A. Laksminingrum, "SaaS Platform for Blockchain Based E-Document Authentication applications," in *2022 International Conference on Science and Technology (ICOSTECH)*, 2022, pp. 1–7.