On Featured IOT for Smart public transportation with cloud-based machine learning

Hatma Suryotrisongko[1], Mochamad Nizar Palefi Ma'ady[2], Hari Ginardi[3]

[1,3]Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia, Raya ITS No. 1
[2]Universitas Nahdlatul Ulama Sunan Giri
Bojonegoro, Indonesia, Ahmad Yani No. 10
[2]nizar@unugiri.ac.id

Abstract With rapid development of science and technology and the accelerated pace of urbanization, smart city has been also widely spread across our country. By Internet of Things, we put forward a kind innovative way of real-time air pollution monitoring into public city transportation vehicle. We take information about the surrounding environment through sensors and upload it directly to the internet. This paper fosters the advantages of the cloud-based computing frameworks (such as Microsoft Azure Machine Learning Studio) and presents a practical work-flow well-suited for the standard machine learning tasks

Keywords: Intelligent Transportation; Smart Transportation; Classification Methods; Internet of Things; MAMLS

I. INTRODUCTION

Nowadays, city transport system has become a part of the urban transportation system, which increases the accessibility of public transportation system. City transport network almost covers the entire urban area of city where city transport system operated. Therefore, it constitutes an ideal Internet of Things (IoT) for monitoring air quality in the city. In the present, many of the systems used for collecting air quality readings are based on sensor network nodes distributed in specific local regions (Tapia *et al.,* 2006).

IoT is a concept and a paradigm that considers pervasive presence in the environment of a variety of things/objects that through wireless and wired connections and unique addressing schemes are able to interact with each other and cooperate with other things/objects to create new applications, services and reach common goals (Ibrahim *et al.,* 2015). The IoT applications are boundless, few examples are; smart cities, smart energy and the smart grids, smart transportation and enabling traffic management and control (Wijaya *et al*., 2017).

IoT of air pollution field mainly uses the sensor, radio frequency, communication and other technologies to detect air pollution, and implement automatic collection and control of gas data, production monitoring and management (Yueling & Yanxiao, 2012). IoT of air pollution field is a unified platform which achieves the common needs of gas production, but due to the different air pollution field business, there are some personalized requirements in the expansion of the IoT field's functions needed some efficient development methods and techniques to achieve.

The proposed cloud-based generalized work-flow allows researchers to focus more on data inferencing than the processing system and data modeling constraints (Qianling *et al.,* 2010). Cloud platforms have been categorized into three categories based on the type of services: Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS). While IaaS allows several virtual systems to operate over a singular hardware infrastructure in an independent manner, SaaS provides installation, management and interoperability of software applications without the knowledge of the hardware infrastructure. The primary advantage of PaaS systems, such as the Microsoft Azure Machine Learning Studio (MAMLS).

With the rapid development of economy, chemical industrial park construction and production activity are increasingly frequent, leading to increasing probability of environmental pollution accidents, especially air pollution accident. Affected by meteorological and geographical conditions, air pollution will be highly clustered in a short time after happening, causing great harm or even extreme destruction to both human and environment (Roychowdhurry & Bihis, 2016). So, it is particularly important to set up a real-time air pollution monitoring system.

Household combustion devices, vehicles and forest fires are common origin of air pollution and noise pollution. Pollutants which are responsible for health concern include particulate matter, carbon monoxide, methane, ozone, nitrogen dioxide and Sulphur dioxide. The literature shows that an IoT application, of which a physical object is embedded with electronics, software, sensors and wireless connectivity to allow monitoring air pollution on real-time (Roychowdury, 2016).

Air pollution and lack of air quality monitoring points represent environmental and technological challenges for cities and environments around the world (Roychowdury, 2016). To face this issue, industry has focused its efforts in finding a versatile technological alternative that allows the improvement of the air quality measuring process and provides reference values in network sites where conventional monitoring fails to cover appropriately.

In this work, we consider a smart city scenario. Smart city is essentially the smart management of city with information technology support, to create a good living and working environment for citizens, improving their living standards and promoting the sustainable development of city (Sutikno *et al.*, 2011*;* De & Leoncini, 2013). Typical characteristics of smart cities are the mobility and the variety of nodes (user mobile devices, sensors, and vehicles), which can be equipped with at least a network interface, and the widespread availability of free wi-fi access points (Sutikno *et al.*, 2011).

We present a smart public transportation which embedding IoT for real-time monitoring air quality as a feature with scalable cloud-computing framework using MAMLS platform to analyse and classify air pollution data set and ensure high classification accuracy.

II. Overview of State-of-The-Art

2.1. Proposed System

The system is comprised of several subsystems as described in Fig. 1. The design process included designing the basic units for sensing the air quality, sending real-time telemetry gases data into cloud-computing platform through Microsoft Azure Portal, and classifying the data set using MAMLS.

Data from gas sensors is sensed with help of inbuilt Arduino and the Raspberry-Pi 3. To present a recharge energy-efficient of headless Raspberry-Pi 3, we make use of solar panel battery bank that can auto-refill the electrical battery from heat of the sun. The data uploaded in Microsoft's Azure Cloud with the help of event hub cloud service and data is stored in the SQL database present in Microsoft Azure Portal. Machine learning service is implemented in machine learning studio which takes data stored in the cloud and perform analysis to classify it into acceptable or affect human health.

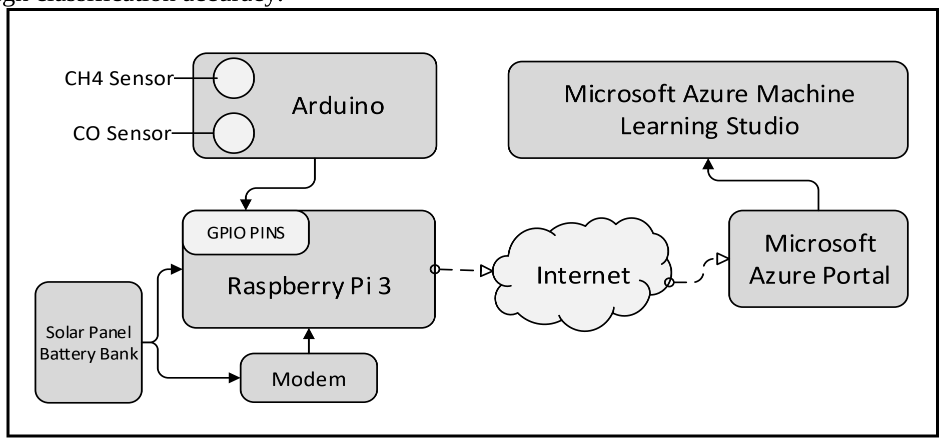

Fig. 1 Architecture of the Proposed System

Fig. 2 A City Transport Vehicle Installation

For some testing, we have deployed mobile sensor node on the city transport vehicles that is actually moving around the city and covering all the major areas of the city. These city transports pass through different areas in which concentration of air pollution is different from each other area. For example, as shown in Fig. 2. For collecting the data of gases in the environment which causes air pollution, these mobile nodes that has been deployed on city transport vehicle collects data when goes in the range of the coordinator node and continuous sensing the concentration. Finally, when the city transport vehicle reaches back to destination, we collected the data collected by routing nodes deployed on city transport service during the whole pathway of the route. Then we analysed the data and collected information and found where the Carbon Monoxide (CO) and Methane (CH4) were in concentration., the node employed on vehicle.

For collecting the data of gases in the environment which causes air pollution, these mobile nodes that has been deployed on city transport vehicle collects data when goes in the range of the coordinator node and continuous sensing the concentration. Finally, when the city transport vehicle reaches back to destination, we collected the data collected by routing nodes deployed on city transport service during the whole pathway of the route. Then we analysed the data and collected information and found where the Carbon Monoxide (CO) and Methane (CH4) were in concentration.

2.2. Materials

For collecting the data of gases in the environment which causes air pollution, these mobile nodes that has been deployed on city transport vehicle collects data when goes in the range of the coordinator node and continuous sensing the concentration. Finally, when the city transport vehicle reaches back to destination, we collected the data collected by routing nodes deployed on city transport service during the whole pathway of the route. Then we analysed the data and collected information and found where the Carbon Monoxide (CO) and Methane (CH4) were in concentration.

The implemented IoT on the proposed system consists of a Raspberry-Pi 3 as single board computer, an Arduino Mega 2560, a modem, a solar panel battery bank, a microSD card, two gas sensors (such as CO and CH4), and several connecting cables as seen in Fig. 3. The measurement system contains the following components: MQ-7 sensor measures the concentration of carbon monoxide (CO) a dangerous, odourless and tasteless gas. The sensor can detect CO gas concentrations anywhere from 20 to 2000 ppm in an easy-to-use method; MQ-2 sensor measures the concentration of methane (CH4), one of the main exposure to dizziness, headaches and a feeling fatigue. The sensor can detect CH4 gas concentrations used to near with smoke fire which urban ambient is from 0 to 10 ppm.

Raspberry-Pi 3 is the newest version of R-Pi that can come with unwired network. It can be useful due to not require any complementary hardware for wi-fi feature. It only need a device for internet transmitter that we provided a modem along with the assembled of proposed IoT.

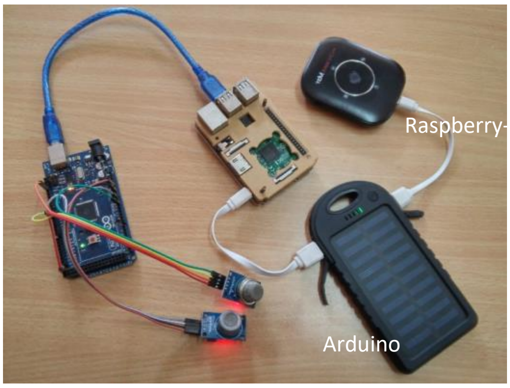

Fig. 3 Internet of Things Assembly

Python programming is used to configure GPIO PINS of the proposed system. Sensor data is stored in the cloud with the help python SQL and event hub API. Generally, socket is used for connection between two platform such as cloud and python. Beside because of the compatibility with programming language of the R-Pi, machine learning service is also deployed and fetched by python script.

PuTTy is a free and open-source terminal emulator, serial console and network file transfer application. It supports several network protocols, including SCP, SSH, Telnet, rlogin, and raw socket connection. The proposed system was running the R-Pi headless. Hence, the software (PuTTy) was used to communicate the R-Pi without the need for hardware peripherals like screen or monitor. It was done by connecting the R-Pi to the internet transmitter and then providing PuTTy with the R-Pi's IP address. By using this method, the R-Pi's power limitations were reduced in order to have a more efficient and reliable system.

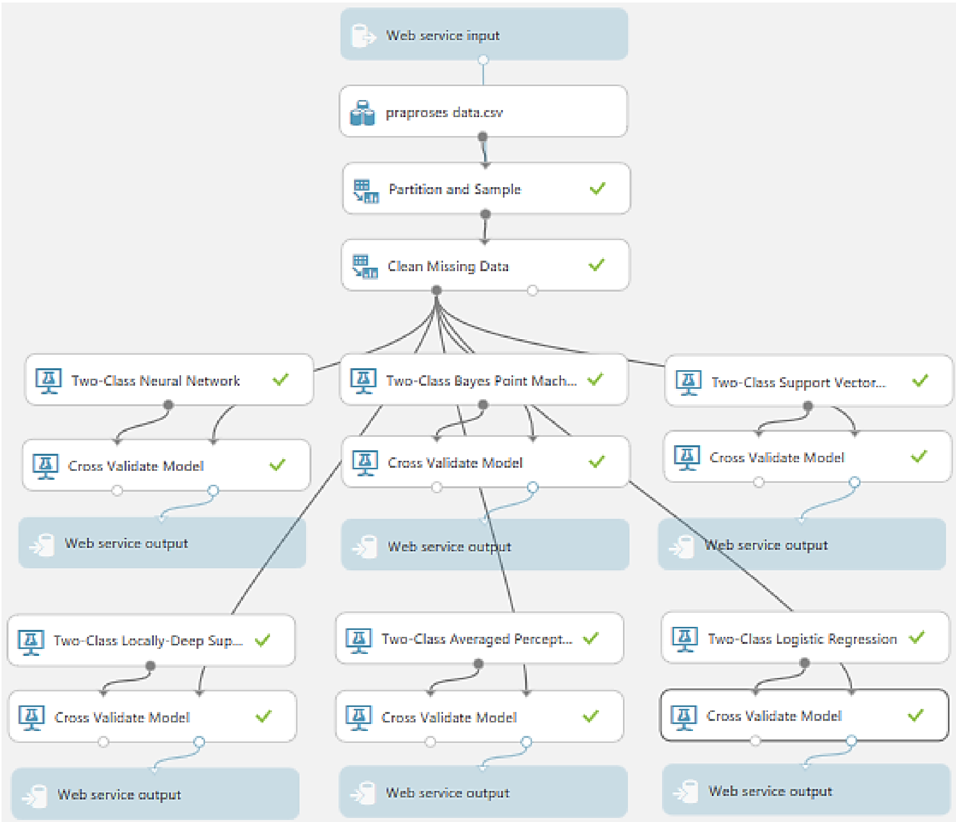

Figure 4. Overall Workflow of The MAMLS Framework

Node-RED is a browser-based visual tool for software development that allows to wire together IoT devices. We divided to deploy a Node-RED application on top of the Node.js runtime in Bluemix. Having considered the complex processing requirements of the sensor management platform, we opted for the PaaS approach. By using the PaaS approach, data collected from air quality sensors can be used to trigger automated actions.

2.3. Cloud-based Machine Learning using MAMLS

The availability of hardware accelerators, such as GPUs, in cloud environments has reduced the computing time for machine learning algorithms on large volumes of data. To establish MAMLS with high accuracy of the classification methods, we set partition and sample, clean missing data, and tested the data set using cross validate model. Cross validate model takes two inputs: a machine learning model and a dataset. To compare the accuracy of cloud-based machine learning of the MAMLS, we used the following six binary classification methods that is two class of: neural network, Bayes point machine, support vector machine, locally-deep support vector machine, averaged perceptron, and logistic regression. We also added web service output to create an API for next developing of the model as shown in Fig. 4.

The advantage of MAMLS will allow contribute to recommendations for building networks to monitor air pollution in order to have a model that can be fed from many points on the planet and can contribute to an information model that collect and analyse large amounts of information in order to create a knowledge base to develop policies and practices according to the objectives of sustainable development.

III. Experiments and Results

The data set contains 10,942 samples with one label and three features including timestamp, CO concentration, and CH4 concentration. The data was obtained randomly in a working-day of city transport along as its pathway route covering the city. This data set has already been cleaned and uploaded to Azure ML. For bench-marking purposes, the proposed generalized flow is tested on six two-class binary classification methods with cross-validate model as the same evaluation technique.

Cross-validate model followed by cleaned data set and classification is performed. We divided the dataset into 5-folds for cross-validation with equal number of rows in each partition. The label of acceptable or affect human health is selected as a stratification key column for sampling. First, each data set is partitioned into training data (30% samples) and testing data (70% samples). Next, the training data set is separated into 5-folds. Once the binary classification is bench-marked, we analyse the performance of data set as seen in Table 1.

Table 1. Performance of binary classification methods on MAMLS

Evaluation Standard	**Neural Network**	**Bayes Point Machine**	**Support Vector Machine**
Accura	89.7%	87.6%	88.1%
cy	87.9%	88.1%	87.9%
Precisi	100%	92.6%	94.8%
on	91.3%	86.9%	88.4%
Recall	97.9%	93.8%	98.5%
F-Score	14.521	5.517 s	4.597 s
Evaluation Standard	**Locally -Deep SVM**	**Averaged Perceptron**	**Logistic Regression**
Accura	90.3%	86.2%	91.3%
cy	%	87.9%	88.7%
Precisi	88.1%	98.6%	100%
on	100%	85.2%	92.1%
Recall	91.6%	96.7%	97.7%
F-Score	99.6%	4.535 s	4.442 s

Here, we observe that for the data set, based on the result, the Logistic Regression can classify obtained gas data with 2% higher classification accuracy and 0.1 second faster elapsed time when compared to the five other binary classification methods. It shows that the Logistic Regression has the best average percentage in evaluation standard that used in cloud-based machine learning. We also concern the differences of queue time such analysing machine learning in cloud-computing platform.

A comprehensive machine learning work-flow that is developed on a shareable cloud-computing platform. Each experiment in the cloud is processed in a queue, and the wait times can vary significantly depending on the service traffic load. Since cloud-based platforms such as the MAMLS are capable of parallel processing of experiment modules, computation time complexities do not pose as bottlenecks.

IV. Conclusion

In the time of science and technology changing rapidly, the trend of smart city cannot be halted. This paper introduces an innovative way for monitoring air quality of city based on the public city transportation system. We focus on the development of vehicle integrated with designed IoT and the advantages of the cloud-based computing framework for the standard machine learning tasks as a feature of smart public transportation. Moreover, an evaluation of proposed cloud-based machine learning has been made, representing high overall classification accuracy and the capability to handle computation time complexities. That way, this paper proposed a good solution to the complexity of air pollution for the development of smarter and sustainable cities, involving their citizens in the environmental protection.

Availability of Data

The Pre-processed Air Pollution data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgements

The authors would like to thank Fathan Nur Hakim from Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, for his contributions on informing about configuration of Raspberry-Pi 3 to the related sensors.

References

1. De, N., & Leoncini, X. (2013). Critical behaviour of the XY-rotor model on regular and small-world networks. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2252-2279.
2. Ibrahim, M., Elgamri, A., Babiker, S., & Mohamed, A. (2015). Internet of Things based Smart Environmental Monitoring using the Raspberry-Pi Computer. IEEE, 159-164.
3. Qianling, L., Bingbing, F., & XingPing, H. (2010). Research and Application of Continuous Integration Based on Hudson. Computer Systems & Applications.
4. Roychowdhury, S. (2016). Classification of Large-Scale Fundus Image Data Sets: A Cloud-Computing Framework. IEEE, 3256-3259.

5. Roychowdhurry, S., & Bihis, M. (2016). AG-MIC: Azure-Based Generalizes Flow for Medical Image Classification. IEEE Access, 5243-5257.
6. Sutikno, T., Facta, M., & Markadeh, G. (2011). Progress in Artificial Intelligence Techniques: from Brain to Emotion. TELKOMNIKA, 201-202.
7. Tapia, E. M., Intille, S. S., Lopez, L., & Larson, K. (2006). The design of a portable kit of wireless sensors for naturalistic data collection. In Persavise Computing. Springer Berlin Heidlberg, ch.8, pp. 117-134.
8. Wijaya, I. D., Nurhasan, U., & Barata, M. A. (2017). Implementasi Raspberry Pi untuk Rancang Bangun Sistem Keamanan Pintu Ruang Server dengan Pengenalan Wajah Menggunakan Metode Triangle Face. Jurnal Informatika Polinema, p. 9.
9. Yueling, Z., & Yanxiao, L. (2012). Design of Intelligent Logistics System Based on Cloud Computing and Internet of Things