Real-time indoor air quality (IAQ) monitoring system for smart buildings

Elias Junior Biondo

Dissertation presented to the School of Technology and Management of Bragança to obtain the Master Degree in Industrial Engineering.

Work oriented by:

Prof. Dr. José Luís Sousa de Magalhães Lima

Prof. Dr. Alberto Yoshihiro Nakano

Prof. Thadeu Brito

Bragança

2022

Real-time indoor air quality (IAQ) monitoring system for smart buildings

Elias Junior Biondo

Dissertation presented to the School of Technology and Management of Bragança to obtain the Master Degree in Industrial Engineering.

Work oriented by:

Prof. Dr. José Luís Sousa de Magalhães Lima

Prof. Dr. Alberto Yoshihiro Nakano

Prof. Thadeu Brito

Bragança

2022

Dedication

I dedicate this work to my family, my teachers and my friends.

Acknowledgement

First, I would like to thank my family, especially my mom Teresinha Ivanete de Lara Biondo and my brother Éder Andre Biondo for always believing in me and supporting me every time I decided to leave my comfort zone looking for my dreams. I got to where I am today thanks to your support and love.

"Nobody dies while staying alive in someone's heart" and even after 12 years I feel you more alive than ever, watching over me from wherever you are and seeing everything I have achieved. I am immensely grateful to my father Albertinho José Biondo, for having taught me values that have helped me in any aspect of my life.

I also thank the friends who were with me since the beginning of graduation, especially the friendships I made this last year during the time I spent in Portugal, you were my family during this period and I will be forever grateful.

I would like to thank my supervisors Professor Dr. José Luís Sousa de Magalhães Lima, Professor Dr. Alberto Yoshihiro Nakano and Professor Thadeu Brito for their teachings and for all the support they gave me during this work, it was a pleasure to work with people like you. Thanks also to all professors who were present and contributed in some way to my academic training.

Finally, I would like to thank the Manoel Costa Lima State College, the Federal Technological University of Paraná, Toledo campus and the Polytechnic Institute of Bragança, which are the institutions responsible for my academic and professional training.

Abstract

Indoor air quality (IAQ) is a term describing the air quality of a room, it refers to the health and comfort of the occupants. Normally, people spend around 90% of their time in indoor environments where the concentration of air pollutants, such CO, CO_2, VOCs, SO_2, O_3 and NO_x, may be two to five times — and occasionally, more than 100 times — higher than outdoor levels. According to the World Health Organization (WHO), the indoor air pollution is responsible for the deaths of 3.8 million people annually. It has been indicated that IAQ in residential areas or buildings is significantly affected by three primary factors: (i) Outdoor air quality, (ii) human activity in buildings, and (iii) building and construction materials, equipment, and furniture. In this contest, this work consist in a real time IAQ system to monitoring and control thermal comfort and gas concentration. The system has a data acquisition stage, where the data is measured by a set of sensors and then stored on InfluxDB database and displayed in Grafana. To track the behavior of the measured parameters, two machine learning algorithms are developed, a mathematical model linear regression, and an artificial intelligence model neural network. In a test made to see how precise were the prediction of the two models, linear regression model performed better then neural network, presenting cases of up to 99.7% and 98.1% of score prediction, respectively. After that, a test with smoke was done to validate the models where the results shows that both learning models can detect adverse cases. Finally, prediction data are storage on InfluxDB and displayed on Grafana to monitoring in real-time measured data and prediction data.

Keywords: Indoor Air Quality, Monitoring System, Machine Learning, Artificial Intelligence.

Contents

List of Tables

List of Figures

Acronyms

AI Artificial Intelligence.

ALRI Acute Lower Respiratory Infection.

ANN Artificial Neural Network.

BPNN Back Propagation Neural Network.

C_2H_6OH Ethanol.

C_6H_6 Benzene.

CH_2O Formaldehyde/Methanol.

CO Carbon Monoxide.

CO_2 Carbon Dioxide.

COPD Chronic Obstructive Pulmonary Disease.

EPA Environmental Protection Agency.

HVAC Heating, Ventilation and Air Conditioning.

I^2C Inter Integrated Circuit.

I^2S Inter-IC Sount.

IAP Indoor Air Pollution.

IAQ Indoor Air Quality.

IAQD Indoor Air Quality Detector.

IDE Integrated Development Environment.

IoT Internet of Things.

LR Linear Regression.

LSTM Long-Term and Short-Term memory.

M2M machine to machine.

ML Machine Learning.

MLR Multiple Linear Regression.

Mod1 Module 1.

Mod2 Module 2.

Mod3 Module 3.

MQTT Message Queuing Telemetry Transport.

NH_3 Ammonia.

NO_2 Nitrogen Dioxide.

NO_x Nitrogen Oxides.

NoSQL Not Only Structured Query Language.

O_3 Ozone.

PM Particulate Matter.

RR Related Risk.

SBS Sick Building Syndrome.

SCL Clock Signal.

SDA Send and Receive Data.

SnO_2 Tin Dioxide.

SO_2 Sulfur Dioxide.

SPI Serial Peripheral Interface.

SQL Structured Query Language.

TSDB Time Series Data Base.

TSR Time Series Regression.

URI Upper Respiratory Infection.

VOC Volatile Organic Compounds.

WHO World Health Organization.

Chapter 1

Introduction

As most people spend around 90% of their time, mainly at home or in the workplace, indoor environment conditions contribute greatly to human well-being [1]. According to the World Health Organization (WHO), the indoor air pollution (IAP) is responsible for the deaths of 3.8 million people annually [2]. Harmful pollutants inside buildings include, but is not limited to, carbon monoxide (CO), carbon dioxide (CO_2), volatile organic compounds (VOC), particulate matter (PM), aerosols and biological pollutants [3]. Therefore, over the past decade, researches on air quality control has begun to shift from outdoor to indoor environments [4]. During the period of the coronavirus pandemic, people have spend more time in isolation where the use of alcohol-based products (such as alcohol gel and disinfectants) has increased significantly since the beginning of the quarantine [5], the exposure of theses substances can cause intoxication, central nervous system depression, nausea, vomiting and shortness of breath [6].

In the early 1970s, scientists started to investigate the cause of complaints in indoor working environments. U.S. Environmental Protection Agency (EPA) studies of human exposure to air pollutants indicated that the concentration of indoor air pollutants may be two to five times — and occasionally, more than 100 times — higher than outdoor levels [7]. Nitrogen oxides (NO_x), VOCs, CO, CO_2, sulfur dioxide (SO_2), ozone (O_3) and PM are examples of indoor air pollutants that have been recognized to have harmful impacts on human health [8].

1.1 Indoor Air Quality

Indoor air quality (IAQ) is a term describing the air quality of a room, it refers to the health and comfort of the occupants. IAQ can be affected by microbial contaminants (mold, fungus) which largely depend on temperature and humidity conditions of a room, gaseous pollutants (for instance CO, CO_2 and VOCs), and dust particles or aerosols. These pollutants can induce adverse health effects to the occupants [9]. In order to protect people from such pollutants, IAQ has emerged and has been developed as a research field [10]. The main parameters for evaluation of IAQ include pollutant concentrations, thermal conditions (temperature, airflow and relative humidity), light, and noise. Thermal conditions are crucial aspects of IAQ because several problems related to poor IAQ can be solved simply by adjusting the relative humidity or the temperature [11].

The Figure 1.1 presents some of the main concepts related to IAQ.

Figure 1.1: Concepts related to IAQ. Source: Adapted from [12]

It has been indicated that IAQ in residential areas or buildings is significantly affected by three primary factors [13][14]: (i) Outdoor air quality, (ii) human activity in buildings, and (iii) building and construction materials, equipment, and furniture. Combustion sources and cooking activities contribute to CO_2, SO_2, CO, nitrogen dioxide (NO_2), and PM emissions into indoor air environments [15][16]. In addition, equipment, such as computers, photocopy machines, printers, and other office machines, emit O_3 and volatile compounds.

1.2 Sick Building Syndrome (SBS)

In 1983, the WHO used the term "Sick Building Syndrome", or SBS, for the first time to describe situations in which building occupants experience harmful effects to the health that appears to be linked to the time spent in a building [17]. It has been reported that symptoms tend to worsen as a function of the exposure time in buildings and can disappear as people spend more time away from the building [18]. According to the WHO, SBS symptoms caused by IAP can be divided into four categories: (i) Mucous-membrane irritation: Eye, throat, and nose irritation; (ii) neurotoxic effects: Headaches, irritability, and fatigue; (iii) asthma and asthma-like symptoms: Chest tightness and wheezing; and (iv) skin irritation and dryness and gastrointestinal problems [19].

1.3 Diseases and infections caused by IAP and low IAQ

The respiratory system is frequently the primary target of IAP effects due to human inhalation. Depending on the place of the affected respiratory tract, acute respiratory infections can be classified into acute lower respiratory infections (ALRIs) and upper respiratory infections (URIs) [20]. URIs are illnesses involving the upper respiratory with common symptoms, such as cough, sinusitis, and otitis media [21], and they are often mild in nature and caused by biological pollutants (viruses, bacteria, fungi, fungal spores,

and mites). Meanwhile, ALRI, an acute infection of the lung, is caused by viruses or bacteria, resulting in lung inflammation [20]. It has been found that IAP increases the risk of childhood ALRI by 78%, which leads to a million deaths in children under 5 years of age every year [22]. Inhaled air pollutants are associated with allergic diseases and pulmonary diseases, such as asthma, atopic dermatitis, and allergic rhinitis [23]. Lung functions can be compromised when exposed to harmful particles like PM and CO [24]. Moreover, the exposed of this substances in the indoor air environment can impact on cardiovascular function [25].

1.4 Objective

In view of the concepts presented about IAQ and the risks that this can be cause to people health if it is not controlled, this work presents the following objectives.

1.4.1 General objectives

This work aims to develop a system capable of monitoring the IAQ, through data stored in a database, and develop a machine learning algorithm to predict the measured parameters and compare to measured data in real-time in order to verify possible anomalies in the concentration of parameters.

1.4.2 Specific objectives

As specific objectives, we have:

- Use sensors to measure parameters that influence the IAQ, such as temperature, humidity, CO, CO2, VOC, NOx and alcohol;
- Create a database from measured parameters;
- Monitor the IAQ system behavior through a monitoring platform;
- Develop a prediction algorithm for machine learning using the created database;

- Train the system using the prediction algorithm to detect patterns that compromise indoor air quality;
- Perform monitoring in real time by comparing the predicted values with the data measured by the sensors.

1.5 Document structure

The document structure is presented as follows:

- Chapter 1 presents what is IAQ and some concepts that influence it, such as IAP and SBS, in addition to the health impacts caused by poor air quality. The objectives of the work are also presented, followed by the structure in which the document is organized.
- Chapter 2 reviews a series of studies about IAQ and monitoring systems to this application. Starting with studies that show the health effects caused by exposition of IAP and low IAQ. Then, some works that develop IAQ monitoring and control systems with IoT. Finally, some researches about artificial intelligence to predict IAQ parameters are presented.
- Chapter 3 presents details about the microcontroller and sensors selected to the system, as well as softwares and tools to data communication and machine learning techniques used to create the predict algorithms.
- Chapter 4 start presenting a problem formulation and a diagram of implementation process of the system. Then, data acquisition and processing are presented using the hardware and software presented in chapter 3. Here, the development of learning algorithms is detailed, as well as the final real-time monitoring application.
- Chapter 5 presents an analyse from the data storage on InfluxDB, the validation of the prediction algorithms and a comparative analysis about LR and ANN performance. Finally, the result of real-time monitoring application is presented.

- Chapter 6 brings the conclusions about the results obtained from the tests achieved and discusses proposals for future work and lines of investigation that could be pursued.

Chapter 2

State of Art

In this chapter, some previous research that brings significant contributions in the area of IAQ will be presented. Starting by showing the main parameters responsible for IAP and poor IAQ, followed by the health effects caused by this factors, such as headaches, nausea, dizziness even more serious situations like chronic obstructive pulmonary disease, asthma, and lung cancer. Next, IoT devices and communication protocols are presented, followed by some applications using this tools to develop monitoring systems. Integration with some kind of control system is also showed with alerts, notifications and target devices such as exhaust fan and ventilation systems. Then, studies on artificial intelligence for prediction focused on application of linear regression and neural networks are presented.

2.1 Health Effects caused by IAP and Poor IAQ

Some of the main sources behind poor air quality are ventilation, building materials, human activities and repeated use of chemical-rich products that are responsible for CO, CO_2, VOCs, O_3, NO_2 and SO_2 concentration in indoors environments [26] [27] [28].

In this context, human beings exhale CO_2 and concentration of this substance in a enclosed space increases as number of people increases causing reduction of cognitive function, headaches and fatigue and, higher concentrations can cause nausea, dizziness, and vomiting. Situations like loss of consciousness can happen too occur at extremely

high concentrations [26]. VOC concentrations in indoor environments are at least 10 times higher than outdoors, regardless of the building location [29]. Most people are not seriously affected by short-term exposure to low concentrations of VOCs, but in cases of long-term exposure, some VOCs are considered to be harmful risks to human health, and potentially causing cancer [30]. Little is known about the effects of inhalation exposure to alcohol-based substances. However, in [31] exposed healthy volunteers to a steam solution of 25% alcohol in water causing coughing and a significant reduction in airflow that persisted for 90 minutes in all subjects. The Table 2.1 presents studies that related the exposure to some of IAQ parameters with health symptoms.

IAQ parameters	**Health symptoms**	**References**
CO_2 and ventilation	Nasal congestion, sore throat and headache.	[32]
VOCs	Respiratory irritation and asthmatic symptoms.	[33]
CO_2, NO_2 and O_3	Asthma, wheeze and breathlessness.	[34]
SO_2, NO_2, O_3 and HCHO	Asthma, wheezing and breathlessness.	[35]
SO_2, NO_2 and O_3	Impaired lung function.	[36]
CO_2 and Temperature	Fatigue, stuffy nose, headache, wheezing, cough with wheezing and fever.	[37]
CO_2, Temperature, RH and Bacteria	Respiratory symptoms and Gastrointestinal symptoms.	[38]

Table 2.1: Relationship between parameters responsible for the IAQ and health symptoms. Source: Adapted from [39].

The air pollution can be the cause and aggravating factor of many respiratory diseases like chronic obstructive pulmonary disease, asthma, and lung cancer [40]. In Hong Kong and Taipei, the increasing of O_3, NO_2, $PM_{2.5}$, and SO_2 levels were associated with increased hospital admission for asthma and pneumonia [41] [42] [43]. A number of studies have consistently documented the association between air pollution and the risk of developing lung cancer; women carry the highest risk, probably due to their increased exposure to indoor air pollution [44] [45]. In addition to respiratory diseases, the exposure to indoor

air pollution has high risk effect on the cardiovascular system too, as well hypertension, coronary heart disease, and stroke [46].

In developed countries, exposure to indoor air pollution from the combustion of solid biofuels is a significant public health hazard predominantly affecting women and small children [47]. In [48] shows that air pollution can affect the developing fetus via maternal exposure, resulting in preterm birth, low birth weight, growth restriction, and potentially adverse cardiovascular and respiratory outcomes. In this regard, a number of epidemiological and clinical studies conducted in low income countries found an association between exposure to indoor air pollution during pregnancy and low birth weight and stillbirth [49] [50] [51]. The results of IAP are reported in terms of 2 million premature deaths per year out of which 54% people die from chronic obstructive pulmonary disease (COPD), 44% due to pneumonia and 2% due to lung cancer [52]. A study made in [53] calculated that the related risk (RR) of low birth weight and stillbirth attributable to indoor air pollution in developing countries was 21% and 26%, respectively. Interventions aimed at reducing exposure to household air pollution will result in an improvement of survival outcomes for all children [54]. Notably, a study evaluating the mortality effects of indoor air pollution estimated that the annual child mortality rate would decrease by 0.1 per 1000 children in the absence of these environmental exposures [55].

2.2 IoT for IAQ Monitoring and Control Systems

The IAQ monitoring requires an adequate selection of sensors and communications protocols. The studies made in [56] show that IAQ is dependent of several thermal comfort parameters, especially temperature and humidity. In the same way, some of the majors gases and substances to IAQ analysis are CO_2, CO, VOC, and NOx [57] [58] [59] [60] [61]. The MQ series sensors are known for cost-effective measurements of gases in the indoor environment [61]. However, MQ135 is highlighted as multi-gas sensor for IAQ measurements being capable to measure many different parameters such as CO_2, benzene (C_6H_6), ethanol (C_2H_5OH), NOx, ammonia (NH_3), and smoke [62] [63]. In addition, the BME680

sensor is widely recommended for measuring VOC, atmospheric pressure, humidity, and temperature [64] [65]. To monitoring formaldehyde (CH_2O) levels the ZE08-CH_2O sensor is used in some researches [66] [67]. When using sensors in monitoring systems, data communication via WiFi is one of the most preferred choice [68] [69] [70]. To transfer recorded data to the dedicated server from the target site several researchers work with Message Queuing Telemetry Transport (MQTT) protocol for its ability to support easy implementation and low power consumption [71].

Several studies about indoor air parameters has being developed in the academic scenery. Saini et al. [72] shows the necessity of developing IAQ monitoring systems for hospitals, schools, offices and homes for enhanced health and well-being presenting the connection between COVID-19 pandemic, public health and indoor air quality while addressing the impacts in people's health and wellness. The authors in [73] and [74] paid attention to monitoring CO_2 concentrations in offices and laboratories, respectively. In [75] an IoT-based indoor air quality monitoring system is proposed to monitor O_3 concentrations in the office. Benammar et al. [76] does an IAQ monitoring system with a set of sensors that collect CO_2, CO, SO_2, Nitrogen Dioxide (NO_2), O_3, temperature, and humidity. In a similar way, the research in [77] the authors monitored the office's CO_2, O_2, VOC concentration, for four consecutive days in summer. Chamseddine et al. [78] monitoring a hospital environment focused some air quality indicators, including CO, CO_2, $PM_{2.5}$, PM_{10} and TVOC. In [79] the system presented is an indoor air quality detector (IAQD) enabling measurement and monitoring of CO_2, PM, temperature and humidity, which was tested in residential buildings. The study in [80] implement a real-time monitoring system to collect the indoor temperature, humidity and CO_2 concentration in a small two-story house.

To realize a properly monitoring it's essential the choice of adequate hardware and software to implement the system. Hapsari et al. [81] and Asthana et al. [82] had used MQ135 sensor with WiFi communication to implement an IAQ monitoring system. The device developed in [83] is a prototype of monitoring system for IAQ designed to use MQTT protocol to send the data, measured by MQ135 and DHT11 sensors, to the cloud.

Muladi et al. [63] determine air quality of a classroom measuring the concentration of O_2, CO, CO_2 and NH_3 gases with MQ135, MG811, MQ7 and ME2O2 sensors. Tong et al. [66] use the formaldehyde sensor ZE08-CH_2O with CC2430 and DTH11 sensors to implement an IAQ system with WiFi communication. In [84] the ZE08-CH_2O complete a low-cost systems of sensors used to monitor the air quality of an electric vehicle cabin. Lasomsri et al. [85] develop a low-cost device with BME680 and CCS811 sensors to monitoring some IAQ parameters in a hospital, such temperature, humidit, CO_2 and VOC. In the same way Kadir et al. [86] developed a low-cost IoT device with BME680 and CCS811 sensors and ESP32 microcontroller equipped with wireless WiFi communication systems that monitors the temperature, humidity, equivalent dioxide carbon (eCO_2), and TVOC. The collected data is stored on the database platform InfluxDB. The microcontroller ESP32 have been used in a lot of researches to IAQ Monitoring, making available some useful functionalities, such WiFi and Bluetooth communication, in addition to low-power consumption [87] [88] [89]. Agbulu et al. [90] use an ESP32 with a set of sensors in an ultra-low power IoT system for IAQ monitoring. The study exploits Deep-sleep, and Modem-sleep configurable modes of ESP32 to complement the Active-mode and aid optimal power-saving.

There are many methods to implement an IAQ control system according to the parameters being monitored. The study presented in [91] presents a solution based on ESP32 microcontroller, in an IAQ measuring system that envisions to reduce the risk of asthma attacks and Chronic Obstructive Pulmonary Diseases (COPD) by monitoring respiratory distress triggering factors. A wireless sensor network (including the MQ135) with multiple measuring nodes, a real-time database, a set of actuators and a web application for the end user was considered to providing different warning notifications when air conditions are critical to the user. Yang et al. [68] shows a system operates over a WiFi wireless network utilising the MQTT protocol to IAQ monitoring as well as controlling an air purifier to regulate the particulate matters concentration, which could be incorporated into such a smart building structure. The work developed in [92] presents a system that can monitor and control the air condition using an exhaust fan, that can be applied to improve the performance and stability of the system. These works used the IoT concept

in conducting real-time monitoring of CO_2 and PM_{10}. The proposed system has fuzzy controls that could adjust the power supply of the exhaust fan automatically depends on the concentration of each pollutant. Due to learning application using fuzzy control, the system was able to remain stable for a longer time, spending less energy to activate the exhaust fan. Mossolly et al. [93] compared some control comfort and indoor air quality strategies in order to increase the energy performance in an academic building, geting 30.4% energy saving were achieved during the summer season of four months.

2.3 Artificial Intelligence to Prediction in Monitoring Systems

Artificial intelligence (AI) can be described as an area of computer science concerned with the modeling of human intelligence in computers [94]. On the other hand, machine learning (ML) is a branch of AI technology that involves the application of statistical methods in enabling computers to learn and process data [95]. Several ML algorithms and techniques are used to prediction in monitoring systems models, the studies presented below will show some of this applications. Ayele and Mehta in [96] proposed an IAQ monitoring and prediction system. The researchers used an MQ135 sensor for gas concentration measurement in the indoor environment. The DHT11 sensor was used to measure temperature and humidity parameters. Saini et al. describes in [97] the functionality of the IAQ monitoring systems for measuring PM_{10} , $PM_{2.5}$, CO_2 , VOC, temperature, and humidity parameters. The authors propose an automated prediction system for PM_{10} using ML. The results show that the proposed system is efficient to support building occupants to prevent critical consequences associated with poor air quality. Similarly, in [98] the study presents an IAP monitoring system that can work as an integral part of a smart home, using a ML algorithm to predicting PM_{10} levels the system presented around 97% of overall accuracy on the tests. Meris et al. [99] presents a device for monitoring IAQ parameters and notifying the user of potential hazards at the place where the device is turned on.

The system also utilize a supervised machine learning to predict new data as accurate as possible. The research developed in [100] shows an indoor CO_2 prediction algorithm to control ventilation and maintain the necessary indoor air quality and reduce energy consumption. The conduction of the ventilation system is adjusted over time based on the signals of the internal CO_2 concentration. Aswani et al. [101] developed a work dealing with a model predictive control of the Heating, Ventilation and Air Conditioning (HVAC) systems able to control the indoor temperature of a room of a computer laboratory in the University of Berkeley.

Linear regression (LR) is an ML approach that relates dependent variable and several independent variables. Because of their ease of use, this technique models have been used in many predict application [102]. Ghosal et al. [103] develop a multiple regression and linear regression algorithms to monitoring and predict deaths count in India caused by COVID-19, using a week death count as input the linear regression model the authors predicted the next week deaths. Catalina et al. [104] developed regression models to predict monthly heating demand for residential buildings, the proposed models showed promising features to be easy and efficient forecast tools for calculating heating demand of residential buildings.

Artificial neural network (ANN) is a nonlinear statistical learning technique inspired by biological neural networks. It is used as a random function approximation tool as the complex relationships between inputs and outputs can be modelled [102]. Nakhi et al. [105] applied ANN to predict the cooling load for commercial buildings and show that a well designed ANN model is able to this prediction only based on external temperature. In [106] Ekici and Aksoy used Back Propagation Neural Network (BPNN), a type of ANN, to predict the heating energy requirements of three different buildings and proved the reliability and accuracy of BPNN to prediction. Mena et al. [107] developed a short-term predictive neural network model to predict the electricity demand of a bioclimatic building. Similarly, the study develop in [108] presents an ANN model to predict hourly electricity consumption of an institutional building. Both studies proved that ANN models can predict building energy consumption fast and accurate. Meanwhile, some researchers

compared ANN with other AI based prediction methods. Farzana et al. [109] used both regression and ANN method to predict annual urban residential buildings energy consumption. In the same way, Zhang et al. [110] applied three regression models and one ANN model to predict HVAC hot water energy consumption. Both researches indicated that ANNs can perform better than regression methods for short-term forecasting. In [111] an ANN model was developed to predict IAQ using six indoor air pollutants and three indoor comfort variables were used as input variables and the occupant symptom metric as output variable. As a result, the constructed network shows a better performance than the multiple linear regression analysis. Ren et al. [112] presents a ventilation control system using a fast prediction model based ANN model, resulting in a favorable performance with energy conservation and indoor pollutant level (e.g., CO_2) respectively decreasing up to 43.8% and 28%. Mumtaz et al. [113] proposes a solution for monitoring and forecasting IAQ parameters, such as CO, NO_2, CO_2, $PM_{2.5}$, temperature and humidity. The system reports real-time air conditions to a mobile app over WiFi and generates alerts upon detection of air quality anomalies. Using a neural network algorithm and applying the long-term and short-term memory (LSTM) model, the system was able to predict air pollutant concentrations with an accuracy of about 99%. In [114] the authors introduce an indoor air quality monitoring system deployed in four Microsoft offices in China. The system also integrates outdoor air quality information with indoor measurements to adaptively control HVAC settings with a view on optimizing the energy efficiency and air quality conservation. Using a neural network-based approach, the time period that an HVAC needs to reduce the concentration of indoor $PM_{2.5}$ into a healthy range is predicted. Zhang et al. [115] shows a study where the indoor $PM_{2.5}$, PM_{10}, and NO_2 concentrations were predicted using multiple linear regression (MLR), time series regression (TSR), and ANN models. The ANN model presented the best performance and could predict the target air pollutants at 10 minutes intervals of the studied building with 90% accuracy levels.

Chapter 3

Developed System Components: Hardwares and Softwares

This chapter presents the tools and devices used in this work, pointing some characteristics and functionalities about all elements of the system. Starting with the hardwares, we present the microcontroller and the set of sensors responsible to measure the IAQ parameters. Then, we present the softwares, such as communication protocols and platforms, to monitoring and data storage. Finally, machine learning techniques to control the system are presented.

3.1 Microcontroller and sensors

Here is presented the hardware of the system, comprised by the microcontroller and the sensors, responsible to measure and collect the data.

3.1.1 ESP32

The microcontroller ESP32 in Figure 3.1 is a single 2.4 GHz WiFi-and-Bluetooth combo chip designed for mobile, wearable electronics, and IoT applications. There are two CPU cores that can be individually controlled, and the CPU clock frequency is adjustable from

80 MHz to 240 MHz. [116].

Figure 3.1: Microcontroller ESP32. Source: Mouser Electronics [117], 2022.

The chip of this module has a low-power co-processor that can be used instead of the CPU to save power while performing tasks that do not require much computing power, such as monitoring of peripherals. ESP32 integrates a rich set of peripherals, ranging from capacitive touch sensors, Hall sensors, SD card interface, Ethernet, high-speed Serial Peripheral Interface (SPI), Inter-IC Sound (I^2S) and Inter Integrated Circuit (I^2C) [118].

WiFi integration allows for a large physical range and direct connection to the Internet via a router. The sleep current of the ESP32 chip is less than 5 μA, making it suitable for battery powered and wearable electronics applications. The module supports a data rate of up to 150 Mbps, and 20 dBm output power at the antenna to ensure the widest physical range [118].

3.1.2 BME680

BME680 presented in Figure 3.2 is a digital 4-in-1 sensor with gas, humidity, pressure and temperature measurement based on proven sensing principles [119]. This sensor is widely used in IAQ, IoT, home automation, control and monitoring systems and weather forecast applications and supports the I^2C and SPI digital interfaces, where it acts as a slave for both protocols. The interface used in this work is the I^2C interface, as it supports Standard, Fast and High Speed modes.

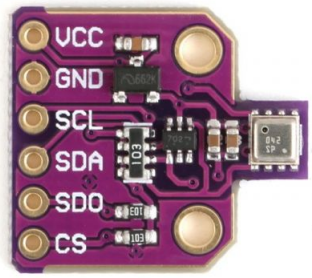

Figure 3.2: BME680 Module Sensor. Source: SATKIT [120], 2022.

This device is a metal oxide-based sensor that detects VOCs by adsorption (and subsequent oxidation/reduction) on its sensitive layer. Thus, BME680 reacts to most volatile compounds polluting indoor air. In contrast to sensors selectivity for one specific component, this sensor is capable of measuring the sum of VOCs/contaminants in the surrounding air. This enables BME680 to detect e.g. outgassing from paint, furniture and garbage, high VOC levels due to cooking, food consumption, exhaled breath and sweating [119]. The signal provided by the BME680 sends resistance values inversely proportional to the VOC concentrations present in the environment, the higher the VOCs concentration, the lower the resistance value and vice versa [119].

3.1.3 MQ135

The MQ135 gas sensor in Figure 3.3 is a low cost sensor with tin dioxide (SnO_2) as a sensitive material, which has lower conductivity in clean air. When there is target polluting gas, the sensor conductivity increases proportional to gas concentration. This sensor is sensitive to some gases, such as NH_3, NO_x, alcohol, C_6H_6 and CO_2, and is used in air quality control equipment for buildings [121].

Figure 3.3: MQ135 Semiconductor Sensor for Air Quality. Source: Winsen [121], 2022.

3.1.4 ZE08

ZE08 in Figure 3.4 is a general-purpose and miniaturization electrochemical formaldehyde detection module. It utilizes electrochemical principle to detect methanol (CH_2O), and is sensible to alcohol, CO and smoke in air with high selectivity and stability. It has a built-in temperature sensor to make temperature compensation and provide a digital output and analog voltage output at the same time. It is a combination of mature electrochemical detection principle and sophisticated circuit design.

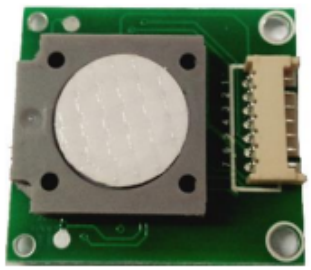

Figure 3.4: Electrochemical CH_2O Detection Module. Source: Winsen [122], 2022.

3.2 Data communication

Here is presented the tools responsible for communicating, storing and displaying the data collected by the sensors.

3.2.1 Arduino IDE

The Arduino Integrated Development Environment (IDE), is an application that can run on different platforms and operating systems, created in the Java programming language, supports programming languages that are based in standard C++ language [123].

Through code developed in the Arduino IDE, the data collected by the sensors are sent to the database through the MQTT communication protocol with the aid of NodeRED.

3.2.2 MQTT

MQTT is an open source "machine to machine" (M2M) protocol widely used in IoT due to its simplicity based on TCP/IP, useful for applications where small amounts of information are sent [124].

The operation of this protocol is based on a client-server messaging service, based on publisher and subscriber communicated through a central server called broker, transmitting information through a payload chain with a hierarchical tree structure, similar to a directory of folders and at the same time to define well the specifications of the message. The parties involved in this process are:

- Broker: central server that distributes and filters information according to a topic. Since the person who publishes the information does not need to know anything about the node that will receive it, furthermore, although the communication of information is usually in real time, the broker has the ability to store messages for clients that are not available at any time;

- Topic: Name of the message that serves as your identifier and where customers will subscribe;

- Publisher: client that sends information to the broker to distribute it to clients who subscribe to it;
- Subscriber: Customers who subscribe to a topic and will receive information about it from the broker;
- Client: The client can be a publisher and a subscriber, receiving information from the medium or performing some interaction function.

The process is divided into four steps which are connection, authentication, communication and termination. The client initiates a connection with the broker, by default on port 1883, and sends it a message for the connection that is serviced by the latter. Once the connection is established, the client publishes messages with the topic and payload, which represent the topic and information. On the other hand, the subscriber must also connect with the broker and subscribe to the topic they want to receive.

3.2.3 Node-RED

Node-RED is an Open Source visual programming environment with the objective of solving the complexity of hardware integration with other services, used to management of real-time data with functions (or nodes) already installed by default. The platform have of a set of libraries available for installation, so if an extra node is needed or for a more specific function, such as InfluxDB or MQTT, these can be installed as long as the corresponding library is available as an open source project [125].

There is also the possibility for the user to create their own nodes, the main way to create a node is through an HTML and JavaScript file with the configuration and programation logic of the node you want to design.

3.2.4 InfluxDB

InfluxDB belongs to Influx Data and is a system for creating and managing databases, it is a free open source software despite having a paid version. It is a software developed for

time databases (TSDB or Time Series Data Base) which, as the name suggests, stores large amounts of time series sorted by time. Unlike relational databases, TSDBs offer faster data storage and processing along with corresponding timestamps, a fact that gained importance with the popularization of IoT and Big Data. On the other hand, InfluxDB goes beyond SQL (Structured Query Language) and NoSQL (Not Only Structured Query Language) schemes because it allows real-time data analysis [126].

3.2.5 Grafana

Grafana is a free monitoring software that allows the representation and visualization of data series from temporary databases. Thus, this tool allows the representation of the chosen data through the database saved by InfluxDB, representing them according to the need within a wide range of options for display and monitoring panels. The software also allows the use of alerts for values that exceed an established range through tools such as Email, Telegram, SMS or Microsoft Teams [127].

3.3 Machine Learning

This section will present two of the most common techniques of machine learning to predict data, being these Linear Regression (LR) and Artificial Neural Network (ANN).

3.3.1 Linear Regression

Linear regression aims to apply a set of assumptions primary regarding linear relationships and numerical techniques to predict an outcome (output) based on one or more predictors (input) with the end goal of getting a mathematical model to do a prediction of outputs, given only the predictor values with some amount of uncertainty [128]. This model is presented being:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_i \cdot x_i + \epsilon, \tag{3.1}$$

where,

- y is the predicted output variable (dependent variable);
- x_i are the predictor variables (independent variables);
- a_0 is the coefficient that intercept or the value of y when each x_i equals zero;
- a_i is the coefficient that change y based on a one unit change in the corresponding x_i;
- ϵ is a random error term associated with the difference between the predicted y value and the actual y value.

3.3.2 Artificial Neural Network

ANN is a popular data-driven method based on an interconnected structure of neurons (connected units or nodes) that work in a complex network of interactions to transmit, collect, and learn information based on a history of the information that has already been collected [129]. It uses a complex combination of weights and functions to convert input variables into predicted variables (outputs) without the need for predefined assumptions regarding the relationship among the variables [130]. An ANN structure generally consists of an input layer, several hidden layers, and an output layer (Figure 3.5).

Each layer consists of a set of neurons (with one or more), where each neuron is connected to previous ones through a specific weight to that path. The two input neurons are fed into the neural network, which are processed and transmitted through two layers of neurons, which are referred to as hidden layers. Figure 3.5 shows two hidden (or deep) layers with each layer containing four neurons (nodes). During the learning stage, a data set of known inputs and outputs is fed to the model. The weights between the neurons are adjusted by iteration until an optimal solution of the outputs is reached. The signal then exits the neural network and is aggregated at the output layer as a single numerical predicted value [128].

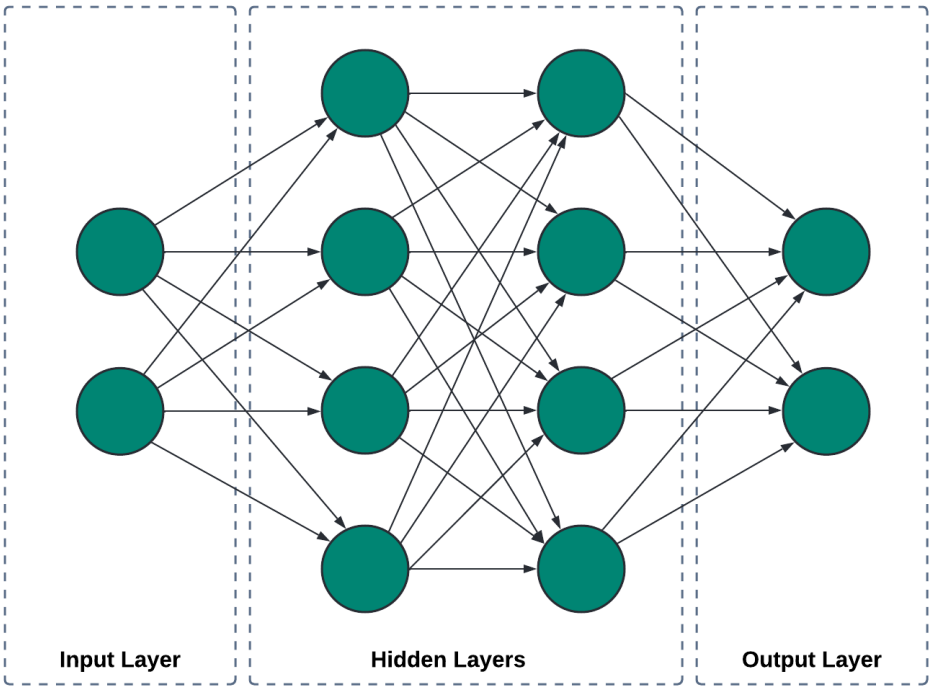

Figure 3.5: Schematic of a four-layer feed-forward artificial neural network.

3.3.3 PyCharm

The linear regression and neural network algorithms will be developed in PyCharm, an IDE used for programming, specifically for the Python programming language [131]. The libraries and tools provided by the software allow you to perform graphical analyses and save data in different formats, such as PNG, TXT and CSV.

Chapter 4

Methodology

This chapter starts presenting the problem formulation, addressing some aspects that involve the IAQ, such as parameters that affect it and the health risks caused by the lack of care with this subject. In sequence, a block diagram with all steps presents on the process to the implement the system develop in this work will be showed.

A step-by-step of data acquisition and data processing is presented. Starting with the hardware architecture of the system showing the wiring diagram and communication interface of sensors with the microcontroller. Then, the process of data transfer and data storage, and finally monitoring and control phases will be presented.

4.1 Problem formulation

IAQ is an area that involves the health and comfort of the indoor environment occupants. People spends most part of your day exposed to some indoor pollutants or low thermal conditions without knowing, since this situation is often not easily noticeable. According the exposure time and the pollutant concentration this exposition can be harmful to health. Saini et al. [72] shows the necessity of developing IAQ monitoring systems for hospitals, schools, offices and homes for enhanced health and well-being and addressing the impacts of this kind of study to people's health and wellness.

Development of this work passed for some phases, starting with understand the problem and the processes that involves an IAQ control and monitoring system through review of studies presented on state of art. Figure 4.1 presents a block diagram of the implementation of the system, starting with the data acquisition phase, that involves the hardware architecture, and data processing with all steps of this phase, that involves the software architecture of the system.

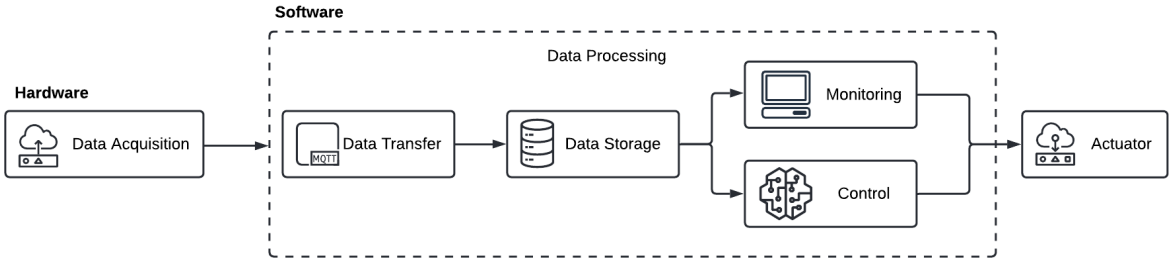

Figure 4.1: Phases of implementation process of an IAQ monitoring and control system.

Next, it is explained the action performed for each phase of the system proposed in Figure 4.1.

4.2 Data acquisition stage

Data acquisition is done by the hardware of the system that is composed by the ESP32 microcontroller and the three sensors, those MQ135, BME680 and ZE08-CH2O. The choice of these devices was based on some IAQ monitoring studies presented in [66] [81] [83] [85] [86], where other sensors, such as DHT11, MQ5, MQ7 and CCS811, were tested before obtaining this configuration. These sensors were chosen due to the fact that they present better stability and precision of the measured parameters, in addition to their low cost. The Figure 4.2 presents the final result of the module with microncontroller and sensors, that make up the data acquisition stage of the system. Three modules were assembled (Module 1, Module 2 and Module 3) where just Module 1 not having the ZE08 sensor due to supplier logistic problems.

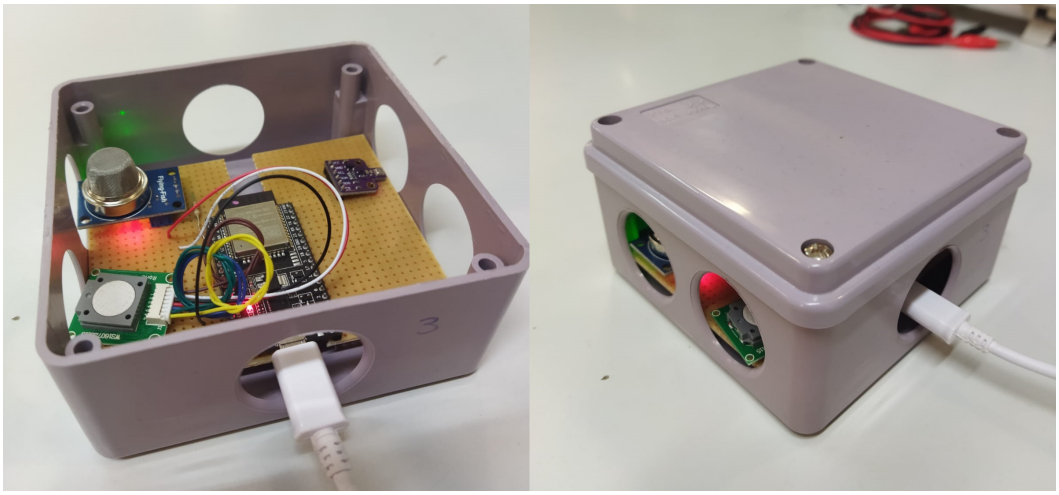

Figure 4.2: Module with microcontroller and sensors used to collect data.

Each sensor has their particularities about PINOUT, interface of communication protocol and power supply. The connection of the sensors with ESP32 is presented in the next sessions.

4.2.1 MQ135

Figure 4.3 show the connection between ESP32 and MQ135 where the Vcc pin is connect to 5 V of the ESP32 board, the ground pin is connect to the ESP32 ground and the analog pin of the MQ135 is connect to the A0 pin of the ESP32 via a resistor.

For this circuit, the A0 port is used, this GPIO allows receiving the analog signal from the sensor without interference due to the use of WiFi. GPIOs require 3.3 V signals (not 5 V tolerant). To solve this, a serial resistor is add between the MQ135 analog pin and the ESP32 A0 pin to protect the GPIOs port from damage.

4.2.2 BME680

BME680 outputs resistance values react according the gas concentration, where the output signal variation is inversely proportional to the gas concentration. The sensor is connected to ESP32 for I^2C communication protocol that uses two wires to share information. One is used for the clock signal (SCL) and the other is used to send and receive data (SDA).

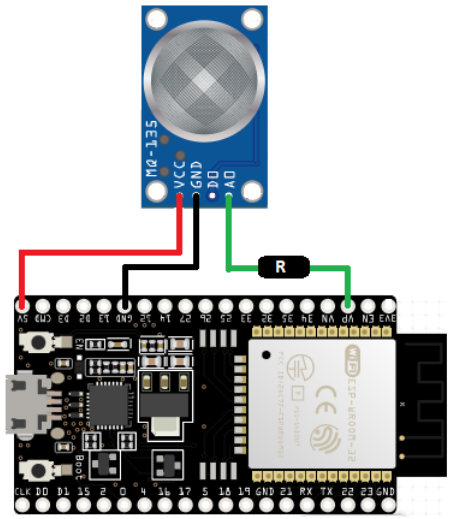

Figure 4.3: Connection between ESP32 and MQ135.

The Figure 4.4 presents how the BME680 is connected to ESP32.

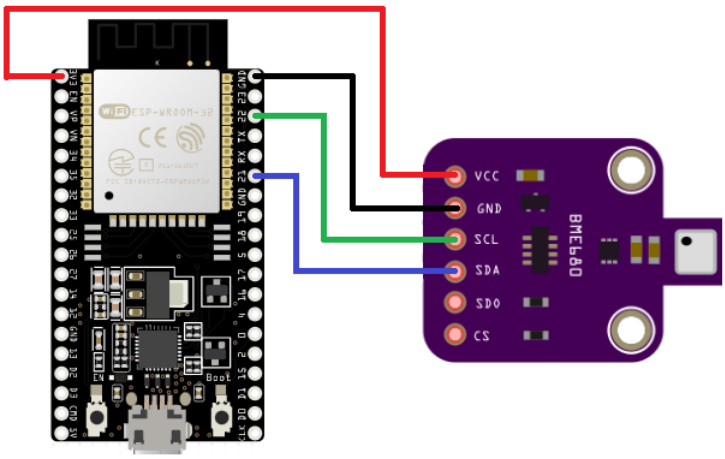

Figure 4.4: Connection between ESP32 and BME680.

The sensor is powered with 3.3 V and the SDA and SCL outputs are connected to GPIO21 and GPIO22, respectively, that is the default to connection I^2C using ESP32 with arduino IDE.

4.2.3 ZE08

This sensor uses only 3 pin to connect with the microcontroller and gives an output signal in a range of 0 to 5 ppm with high sensitivity (of up to 0.01ppm). Figure 4.5 shows the connection.

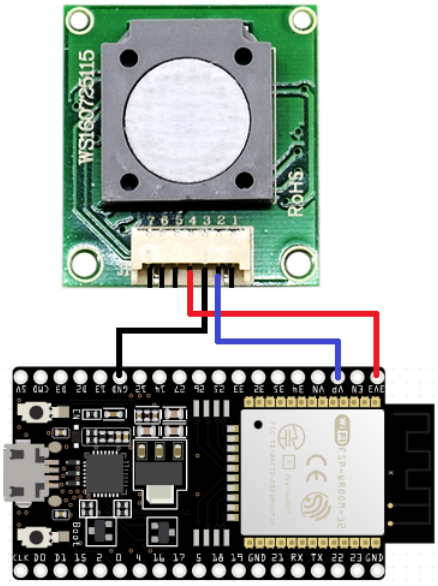

Figure 4.5: Connection between ESP32 and ZE08-CH2O.

Pins 3 and 4 are GND and VCC, respectively. Pin 2 of ZE08 is connected to Vp pin of ESP32, as well as MQ135 this sensor gives measurement through an analog signal.

4.3 Data processing stage

Data processing is done by a set of tools and software (Arduino IDE, MQTT protocol, Node-RED, InfluxDB, Grafana and PyCharm) commonly used for monitoring systems [68] [81] [83] [132] [133] [134]. Figure 4.6 shows the phases of this process.

The data transfer is responsible to receive the data with Arduino IDE and send to influxDB using MQTT protocol through Node-RED platform. Then, stored data are sent

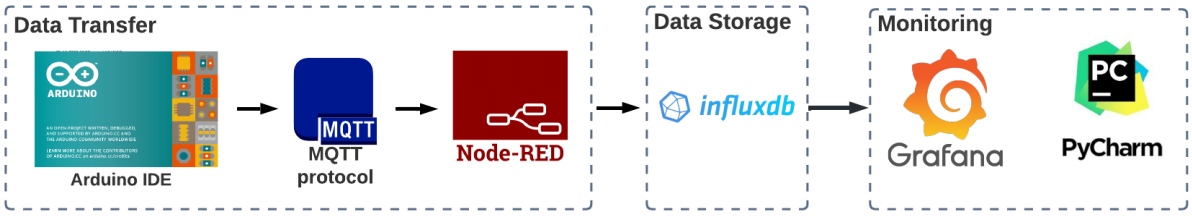

Figure 4.6: Software diagram of the system.

to Grafana to be monitored and accessed by machine learning algorithm developed on Pycharm.

4.3.1 Data Transfer and Data Storage

Figure 4.7 is the flowchart of the Arduino IDE code develop to receive the data read by the sensors and send to MQTT topic.

As shown in chart, the algorithm start initialising some parameter, importing the libraries of the ESP32 and the sensors, setting pins that will receive the data of each sensor and functions to configuring I^2C protocol communication and the functions to connect the microcontroller to WiFi and the MQTT client. Then, the loop starts checking if WiFi and MQTT client are connected, if not, it will call the respective functions to reconnect. Finally, to each parameter measured, a MQTT topic on Node-RED will receive the information and store data on influxDB, as shown in the Figure 4.8.

The node A is an MQTT node responsible to receive data from Arduino IDE to MQTT topics. Then, in node B the data passes through a function responsible for assigning an identifier for each parameter being these: gasBME, altBME, humidBME, ppmMQ135, pressBME, tempBME, ppmZE08. Finally, in node C this information is joined in a vector and sending to influxDB in node D.

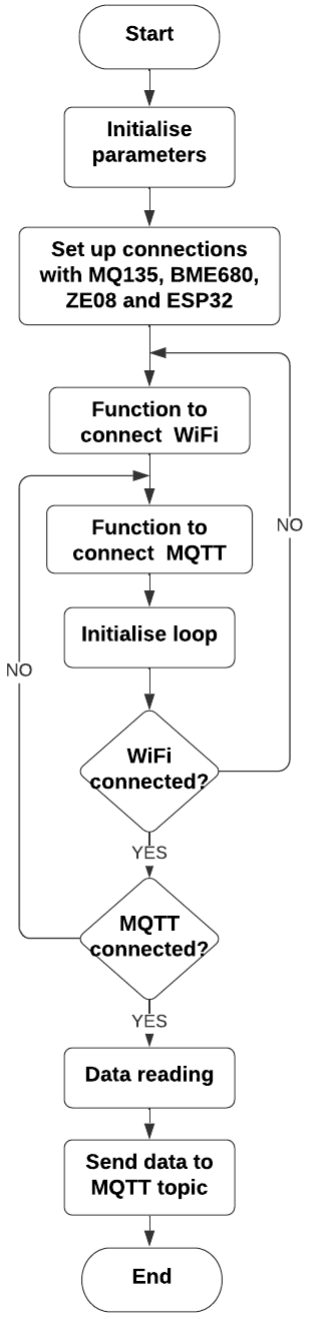

Figure 4.7: Flow chart of the algorithm developed in the Arduino IDE.

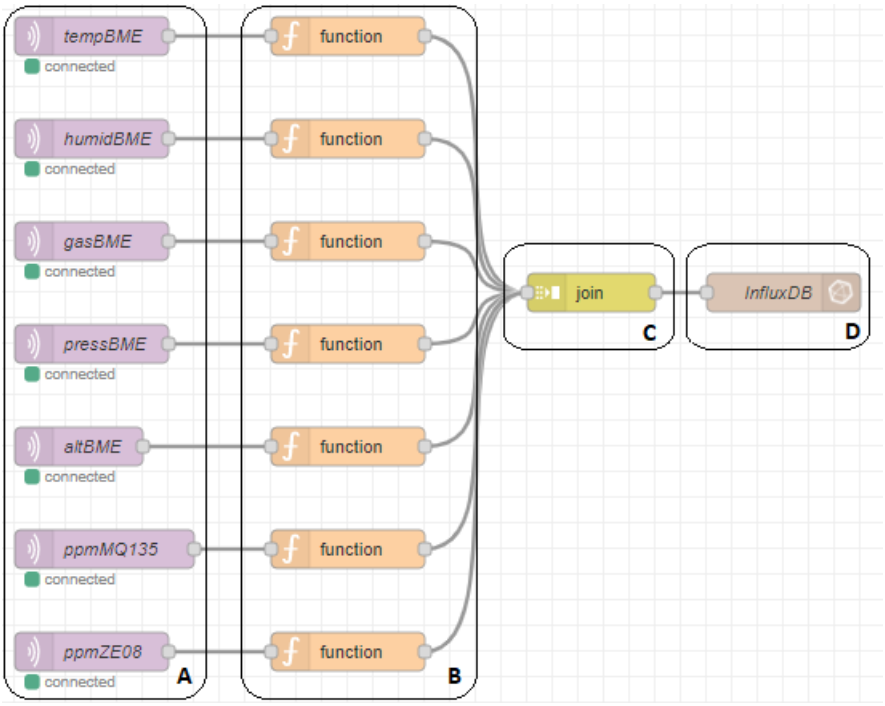

Figure 4.8: Node configuration on Node-RED.

4.3.2 Monitoring

One of the tools being used in monitoring phase is Node-RED. In Figure 4.9, a dashboard was created to show in real time all measured parameters.

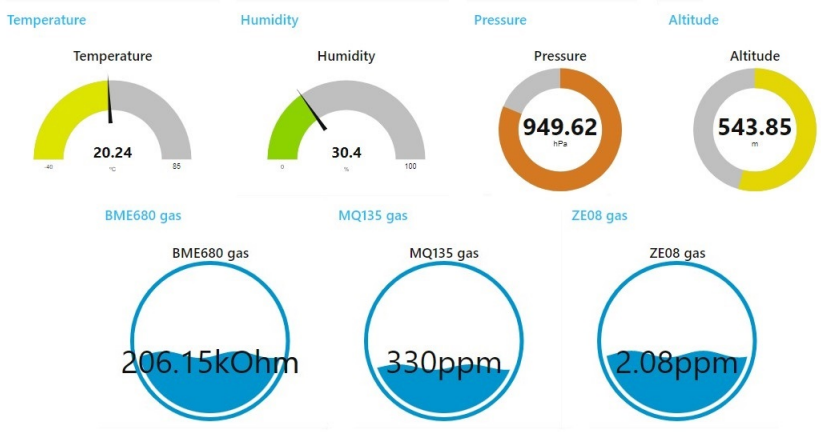

Figure 4.9: Data monitoring in Node-RED dashboard.

All parameters are indicated with their related units and scale according the values indicated in the data-sheet sensors. After store on database, data is displayed in Grafana

where is possible plot graphics and visualise parameters behave throughout every day since the database was created. Figure 4.10 presents a dashboard with measurements of all parameters measured from January 16th to 23th, 2022.

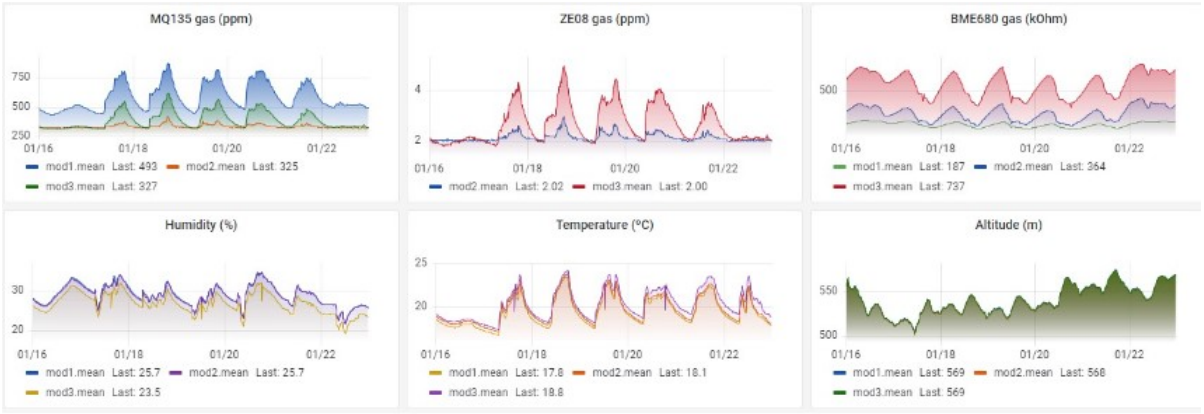

Figure 4.10: Data monitoring in Grafana.

The graphics shows the modules data during one week indicating the last value measured by all sensors (mod1, mod2 and mod3).

Database was also used in a linear regression and a neural network algorithms to predict the values of the measured parameters and detect non-standard behavior. The software used to develop these codes was PyCharm IDE. The step-by-step of the LR and ANN algorithms are presented through a flowchart in Figures 4.11 and 4.13, respectively.

To facilitate the understanding, the explanation of the algorithm will be made for the temperature variable, however this process was carried out for all parameters. The code starts loading data from influxDB corresponding to the last 3 days of temperature reading, where data from last 24 hours will be used for the test model and data from 48 hours before are for training model. Then, a sample of the average temperature is taken every 15 minutes, then it goes through a derivation process generating new sequences of temperatures shifted in relation to each other. To better understand the derivation process, see Figure 4.12 where an example of 8 temperature samples and 3 stages of derivation of this parameter (temp_1, temp_2 and temp_3) is presented.

At each derivation, the temperature values are shifted by a sampling interval ahead

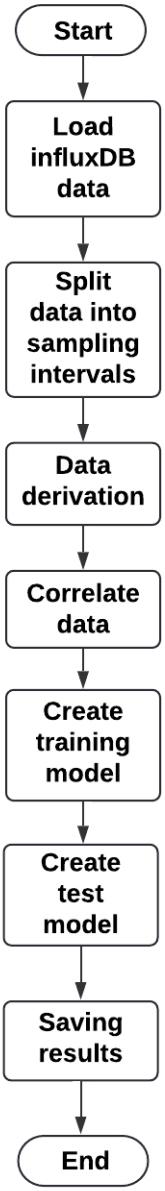

Figure 4.11: Flow chart of linear regression algorithm.

Samples	1	2	3	4	5	6	7	8
temp (°C)	15	16	17	18	19	20	21	22
temp_1 (°C)	0	15	16	17	18	19	20	21
temp_2 (°C)	0	0	15	16	17	18	19	20
temp_3 (°C)	0	0	0	15	16	17	18	19

Figure 4.12: Stage 1 of temperature derivation.

and the sample values corresponding to the previous intervals of each derivation are null, so all the data corresponding to these intervals are taken from the dataset so that it can work only with values non-null, highlighted in Figure 4.12.

Then, forecast models are created based on correlations, made between temperature and its derivatives in order to establish a relationship between the current value of the parameter and its past values, so that it is possible to predict future values and using respectively 90% and 10% from the database for training and testing. In the end, the results from temperature prediction values, assertively coefficient and mean absolute error are saved.

In appendixes A.1 and A.2 a pseudo code from this algorithm tested for different sampling intervals and number of derivations in order to obtain the best possible prediction are presented. The result of this analysis is presented on the next chapter, just like the test done to prove the learning functioning.

The Figure 4.13 shows ANN algorithm flowchart where the initial part of loading influxDB data, sampling and derivation is done as same as LR algorithm. After this, the dataset is divided into training, testing and validation being 80%, 10% and 10% respectively. Then, 2 layers deep was defined where both layers have a width of 50 nodes with the dataset, with the exception of the parameter under analysis, used to create a reusable function. Next, the training loop will train the model with training data of reusable function and evaluate it periodically with the validation data. This validation process helps ensure that the model is not overfitted to the training set data.

Overfitting occurs when the model fails to generalize and make accurate classifications on data it has not been trained on. During training, a validation dataset is used to make sure that the results will be as good as the training, and to verify that the model is not overfitted. Finally, the prediction function will use the reusable inputs already trained to predict the parameters and then results of prediction values, precision and mean error are saved. Figure 4.14 shows a schematic of the final training model of ANN projected on code to predict temperature, the same is done with all parameters.

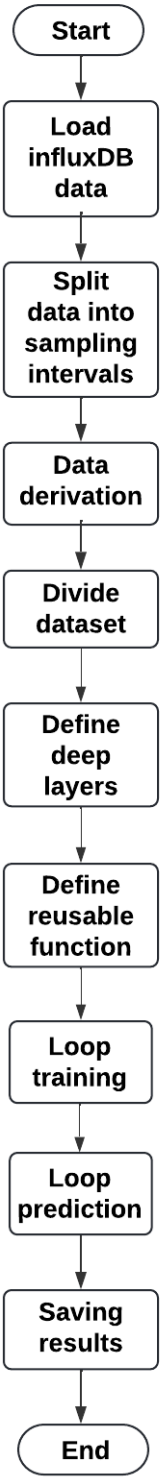

Figure 4.13: Flow chart of neural network algorithm.

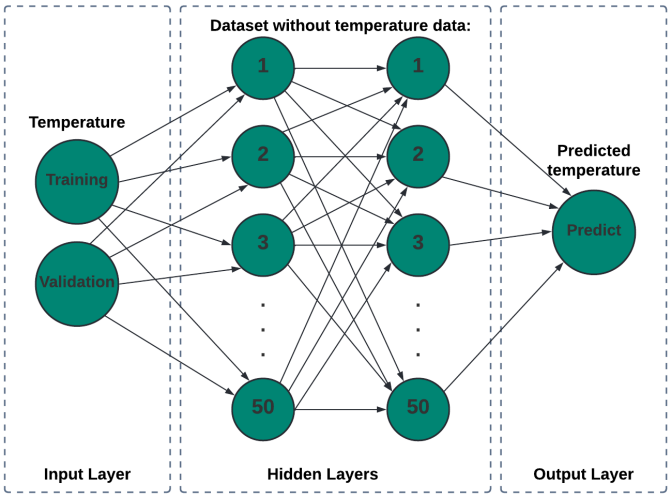

Figure 4.14: Schematic of the neural network training model projected.

During the training phase, a series of trainings are performed, where the reusable function generated in each test is used to feed the next test until the model obtains the best prediction accuracy score.

4.3.3 Real-Time Monitoring

After training the models, the prediction data is stored in InfluxDB through NodeRED and displayed in Grafana in real-time along with the data measured by the sensors. Figure 4.15 shows nodes on NodeRED responsible for this process. In A the node send a command every 2 minutes to compile the algorithm and produce a new prediction that uses a function called datetime.today() to do the process with the actual time. Then, in B the prediction is sent to database through the InfluxDB node.

Real-time monitoring is done by verifying that the measured data is showing values close to the forecast data displayed in Grafana.

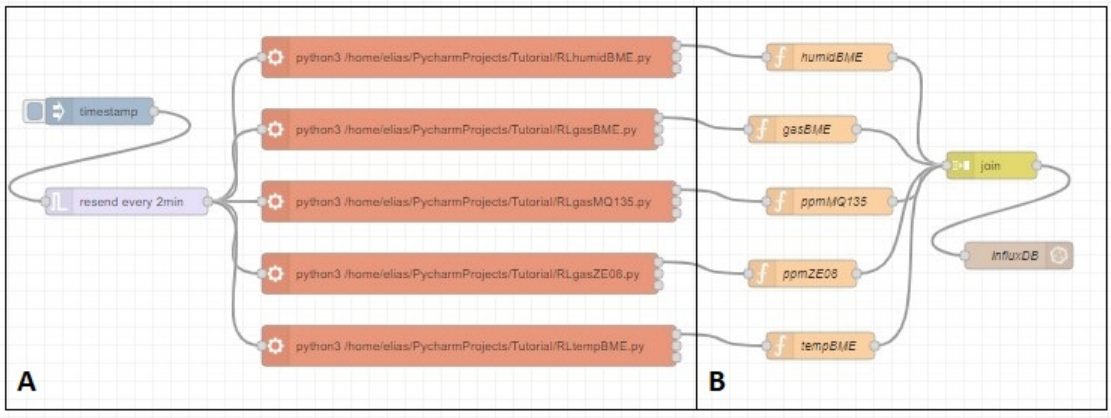

Figure 4.15: NodeRED nodes to storage prediction data in real-time.

Chapter 5

Tests and results analysis

This chapter will present the database collected from 11/23/2021 to 02/16/2022 and an analysis of the data behavior in that period. Next, the process for defining the sampling time and the number of derivation in linear regression algorithm that results in the best prediction is presented. Then, a test with data storage is done to evaluate the accuracy of the prediction and finally a smoke test to validate the functionality of LR and ANN models is done.

5.1 Data behavior analysis

Each module (Figure 4.2) was programmed to collect and store data every two minutes. In Figure 5.1, the database of parameters measured in a laboratory of Institute Polytechnic of Bragança (IPB) is presented graphically, in this period the sensors showed good stability. It is noted that from 12/20/21 to 01/10/22, the concentration of gases showed levels below normal, due to the pause in academic activities at the IPB, which reduced human activities in the laboratory.

Now, to understand the behavior of these parameters on ordinary days of activities at Cedri, data recorded from January 16th to 23th will be presented. Figure 5.2 shows gas concentration measured by MQ135 and ZE08 sensors over the week, where the module 1 (mod1) is presented in blue, module 2 (mod2) is orange and module 3 (mod3) is the

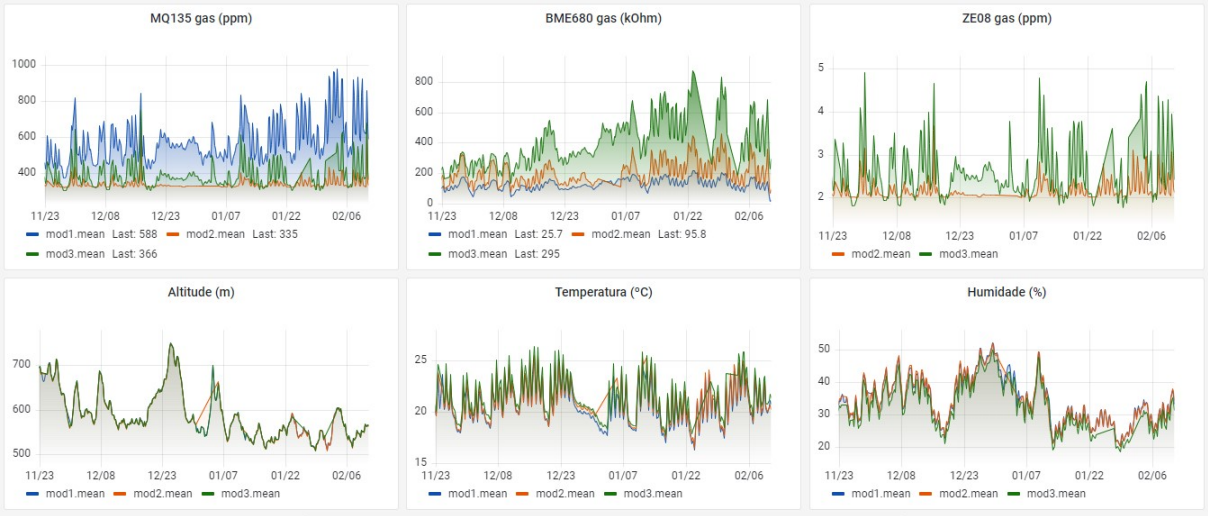

Figure 5.1: Temporal graphs of data stored in the database.

green indicator, in the second graphic only mod1 is not showed since this module does not have the ZE08 sensor.

Starting on Sunday (January 16th) until Saturday (January 22th), the data measured by the sensors showed the same behavior due to some substances, such as CO, alcohol and smoke, which trigger the two devices. From Monday to Friday, gas level increase from 8hr00 morning, where the human activities starts on Cedri, until 18hr00 and then the concentration decrease, due to people starts go way from laboratory, marking the lowest level of gas between 7hr00 and 8hr00 in the morning after passing the period of dawn without human activity in the environment. Weekend has low gas concentration because normally does not have people in laboratory.

Figure 5.3 presents the BME680 reading where the gas concentration showed a behavior similar to the previous cases. However, as the output signal of this sensor is inversely proportional to gas level, in this case the highest values of the output signal were recorded during the dawn period and on weekends, when the lowest levels of pollutant in the environment are presented. The opposite happens in the morning and afternoon periods from Monday to Friday, where the higher the registered gas level, the lower the values of the BME680 output signal.

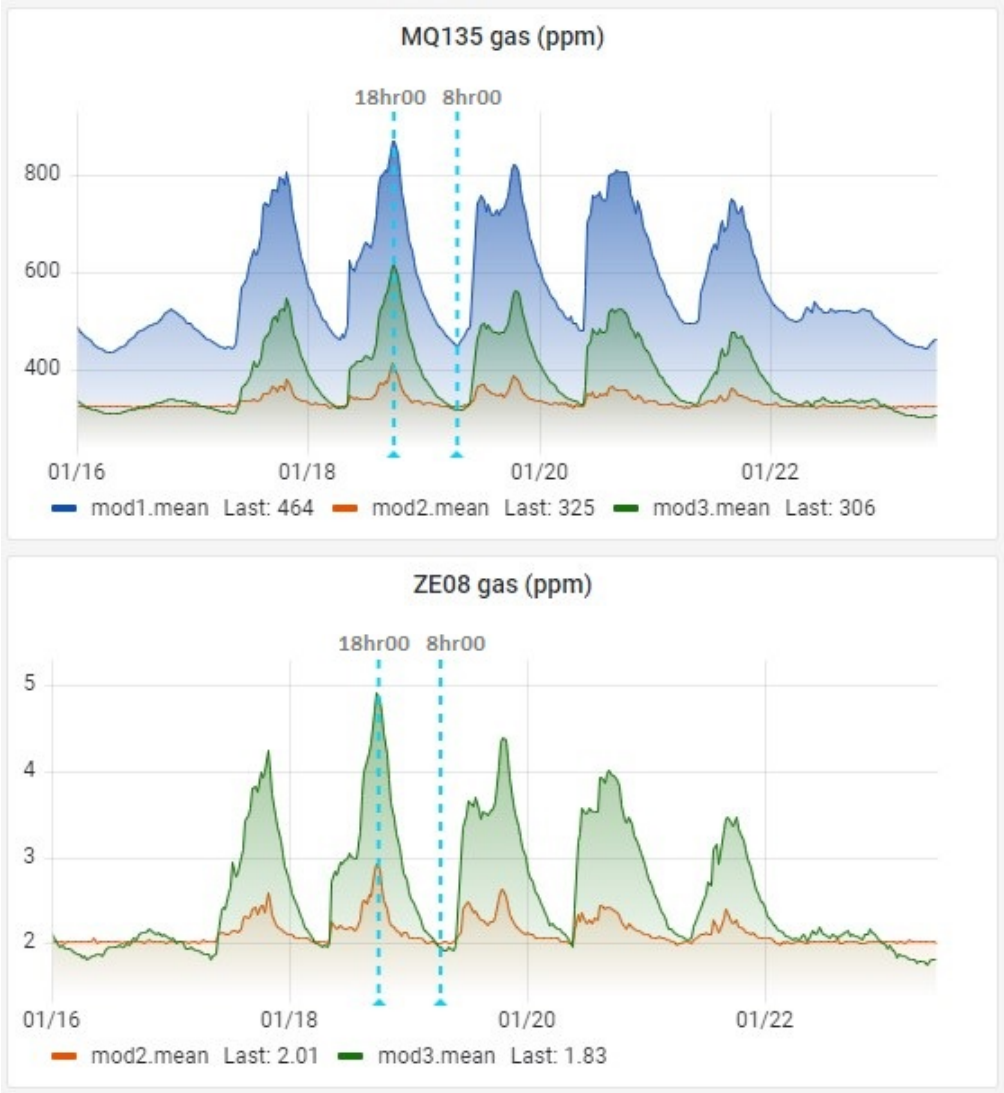

Figure 5.2: Gas measure of MQ135 and ZE08 sensors during a week.

As expected the temperature and humidity levels increases during the day and decrease in the night period. It can be see in Figure 5.4, where this behavior was registered for all week except on Sunday (January 16th) that the temperature scored lower levels than the other days, probably on this day the room heating system was not on.

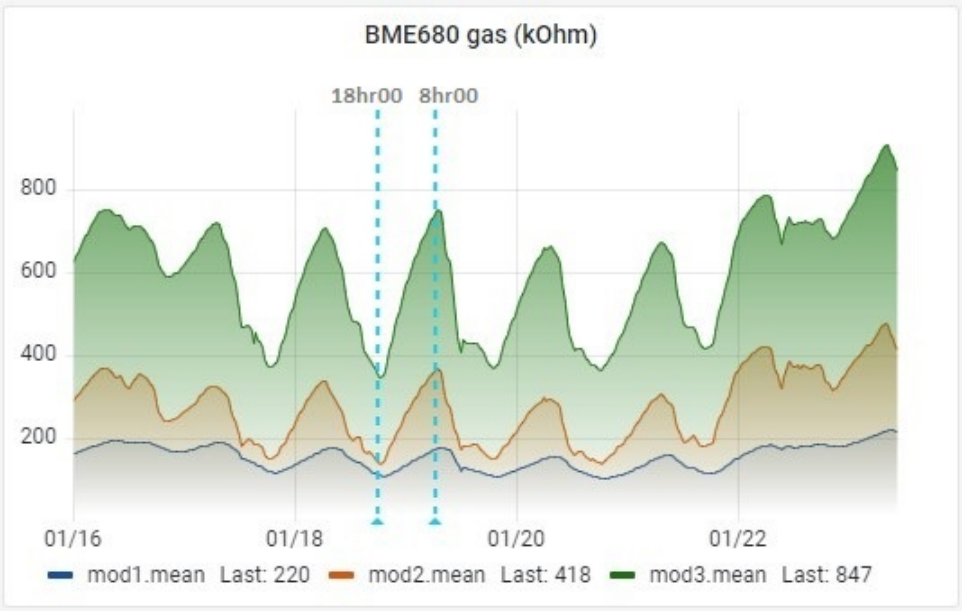

Figure 5.3: Gas measure of BME680 sensor during a week.

5.2 Code analysis

After finishing the LR algorithm, the code was tested to verify result precision from all combinations of the loops of sampling interval and number of derivation to obtain the best possible prediction. Figure 5.5 presents all precision score resulted by a test done to gas concentration measured by MQ135, where this values were summed according to the sampling time and quantity derivation and stored in **Sum Time** and **Sum Deriv**, respectively.

In this case, the best predictions are made using the sampling interval of 1.5 hours (1 hour and 30 minutes) and for two derivations (D2). The same process was done to all others parameters measured for the others sensors and the conditions that satisfy most of the data is 15 minutes sampling and 2 derivation. After summed all Sum Time and Sum Deriv for each case MQ135 gas concentration presents the greater degree of assertiveness. The complete table with all parameters registers are presented in Figure A.3 in Appendix. Then the code was adapted to working with only 15 minutes of sampling time and 2 derivations. Next, the algorithm was tested using data from the same week presented on previous Grafana graphics analysis to prove the precision of the prediction.

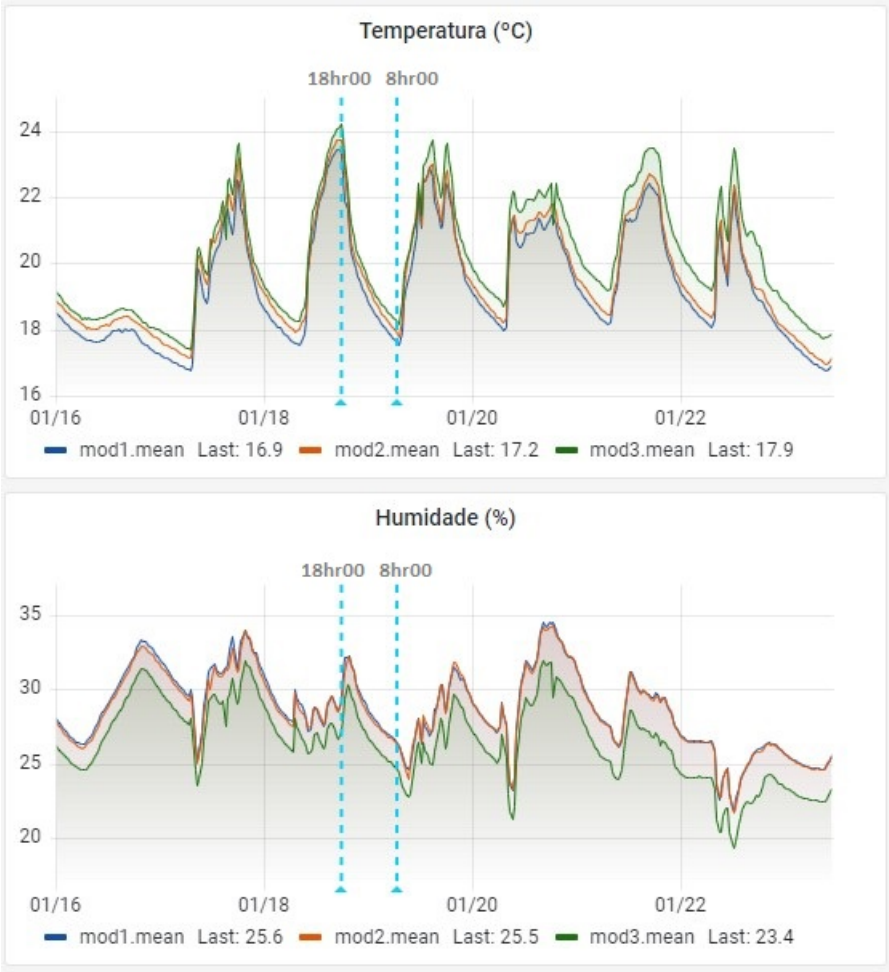

Figure 5.4: Temperature and humidity measure of BME680 sensor during a week.

ppmMQ135					
Sampling time	D1	D2	D3	D4	Sum Time
15min	0.9967	0.9967	0.9968	0.9968	3.987
30min	0.9978	0.9978	0.9918	0.9981	3.985
45min	0.9985	0.9987	0.9986	0.9986	3.994
1H	0.9988	0.9989	0.9987	0.999	3.995
1.25H	0.9992	0.9991	0.999	0.9991	3.996
1.5H	0.9994	0.9994	0.9992	0.9994	3.997
1.75H	0.9995	0.9994	0.9992	0.9984	3.996
2H	0.9994	0.9995	0.9995	0.9773	3.976
Sum Deriv	7.989	7.990	7.983	7.967	31.928

Figure 5.5: A1 results of gas concentration from MQ135.

5.3 Predict test

Using data from January 16th to 23th, the prediction algorithms were tested and the results will be presented below. Figure 5.6 shows LR (A) and ANN (B) forecasts of the gas concentration measured by BME680. For both predictions the red line is the real data used to training the model, the blue line to prediction data resulted from LR model, the green line to ANN model and the grey line represents the error between two values done through:

$$Error = Testdata - Predictiondata. \tag{5.1}$$

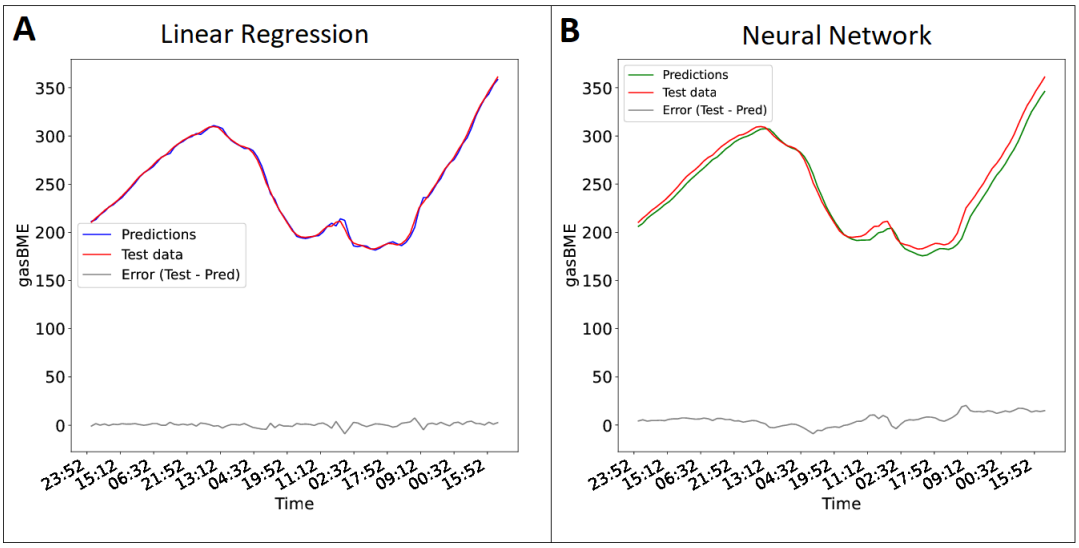

Figure 5.6: Prediction of BME680 gas concentration.

In LR prediction (Figure 5.6.A) the random time showed in x axis is provided from random_state() function used on code that select training data randomly. When analyzing the test and prediction values in both cases the Error value is close to zero for all tested values. The assertively was 99.7% and the mean absolute error to this prediction was about 1.67 kΩ. ANN prediction was similar to LR and scored 98.3% of precision and 7.06 kΩ of mean error.

The Figure 5.7 show the forecast from gas level registered by MQ135. As well as

BME680 gas, LR prediction presented a high assertively degree (95.81%) and about 1.4 ppm of mean error. Looking at the Figure 5.7.B, we see some points where the neural network model presents delays in the prediction in relation to the LR model and because of that this model presented a lower hit rate than the mathematical model, being 81.6% and an average error of 3.11 ppm.

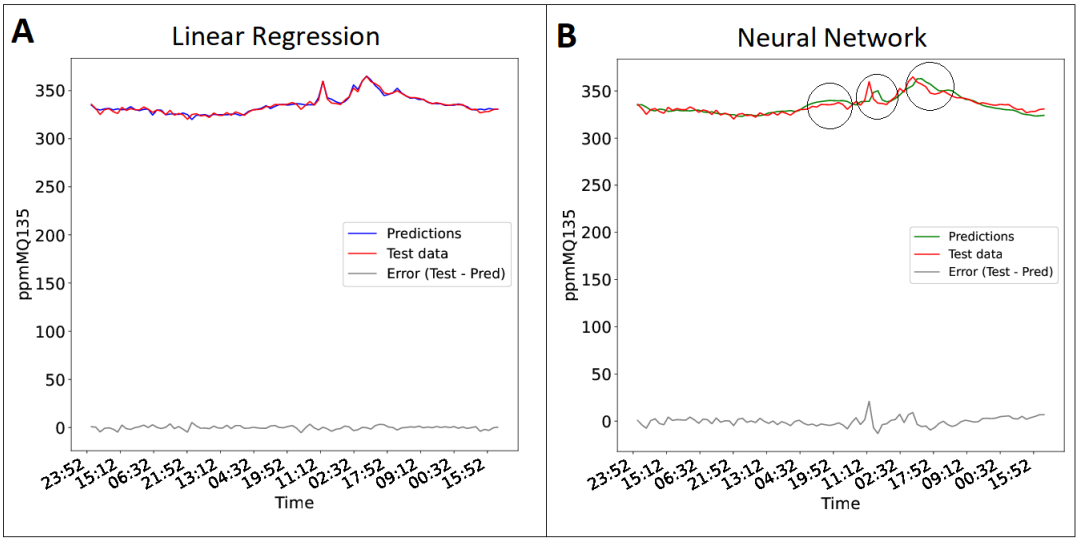

Figure 5.7: Prediction of MQ135 gas concentration.

As well was showed before in 5.1 section, the prediction of ZE08 gas in Figure 5.8 is similar with MQ135 and scored 95.5% of predict precision to LR method and 0.015 ppm of mean error. The prediction on ANN model in Figure 5.8.B presents 81.9% of precision and 0.03 of mean error.

In temperature forecast is notable the difference of precision between LR and ANN models presented on Figure 5.9. The score precision of LR prediction is 99.1% and to ANN is around 82.5%.

The same situation happens to humidity where linear regression done a greater job than neural network to predict this parameter. In Figure 5.10.A the model presents 96.7% of precision and the score in B, presented in the Figure 5.10.B is only 75.9%.

In this test, a dataset was used where the parameters do not present changes in behavior, just to verify the degree of assertiveness of the two machine learning models

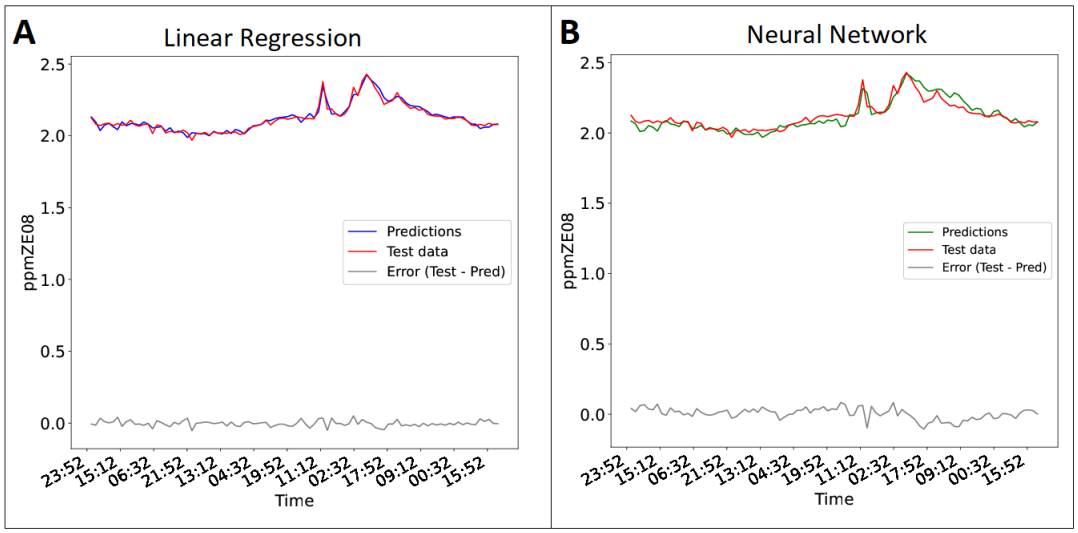

Figure 5.8: ZE08 gas concentration forecast.

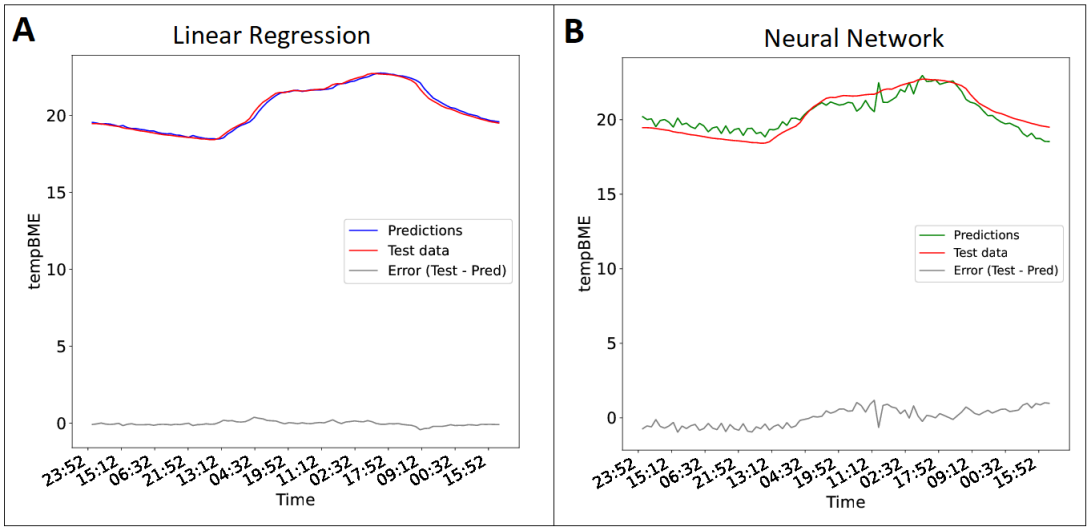

Figure 5.9: Temperature forecast.

developed, in all cases the linear regression model showed a higher degree of precision. to predict the data, another test will now be presented where changes in gas volumes were proved to verify if the algorithms will respond in some way to these changes.

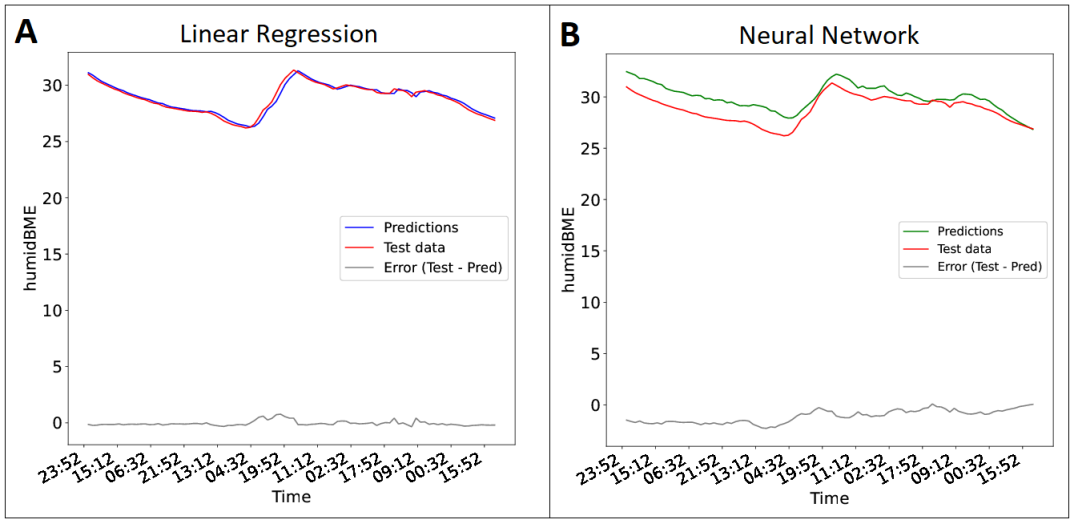

Figure 5.10: Humidity forecast.

5.4 Predict validation

To validate the prediction of the system, it is necessary that the models be tested for adverse values. During the period recorded in the database, there was no abnormal situation in which could be used to prove the operation of the algorithm. For this, a smoke test was carried out on February 10th so that the sensors could read high concentrations of different gases.

5.4.1 Test

In the test, a piece of burnt paper was placed next to the modules and they were covered with a box measuring 60 cm wide, 42 cm deep and 65 cm high. The Figure 5.11 shows the three modules under the box.

The test was carried out for one hour and only the gas levels were changed, while temperature and humidity maintained the same behavior. Figures 5.12 5.13 5.14 shows gas level measured for MQ135, ZE08 and BME680, respectively.

Before start the test the MQ sensor was reading gas levels from around 600 ppm, 300 ppm and 350 ppm to mod1, mod2 and mod3, respectively. Then, the concentration starts

Figure 5.11: Simulation of high gas level with burned paper.

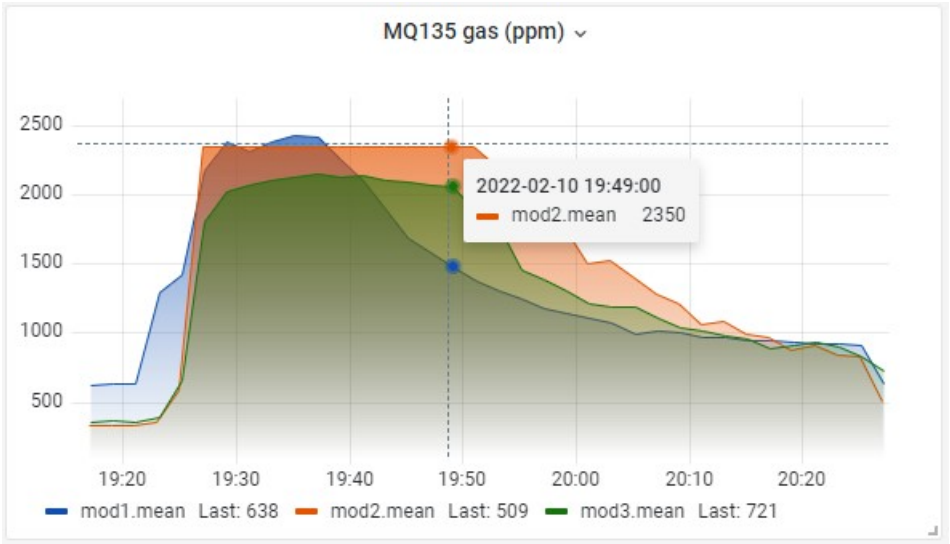

Figure 5.12: Resulting graph of MQ135 gas level for smoke test.

increasing marking values up to 2350 ppm, maintaining this average value for around 30 minutes while there was a concentration of smoke inside the box, as indicated in the figure. As the paper smoke dissipated the gas levels decreased until they returned to normal around 1 hour after the start of the test.

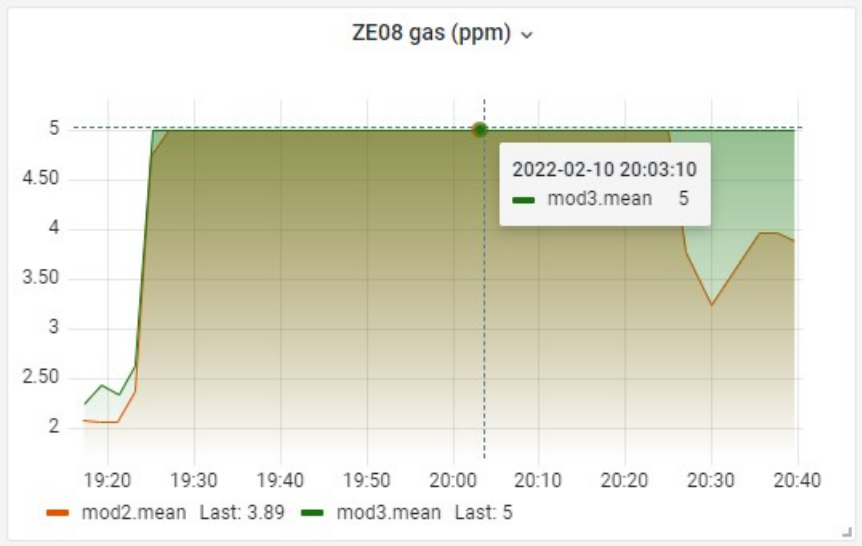

Figure 5.13: Resulting graph of ZE08 gas level for smoke test.

The sensor ZE08 has high sensibility and after start the test both modules that have this sensor registered max level for almost all period of the test.

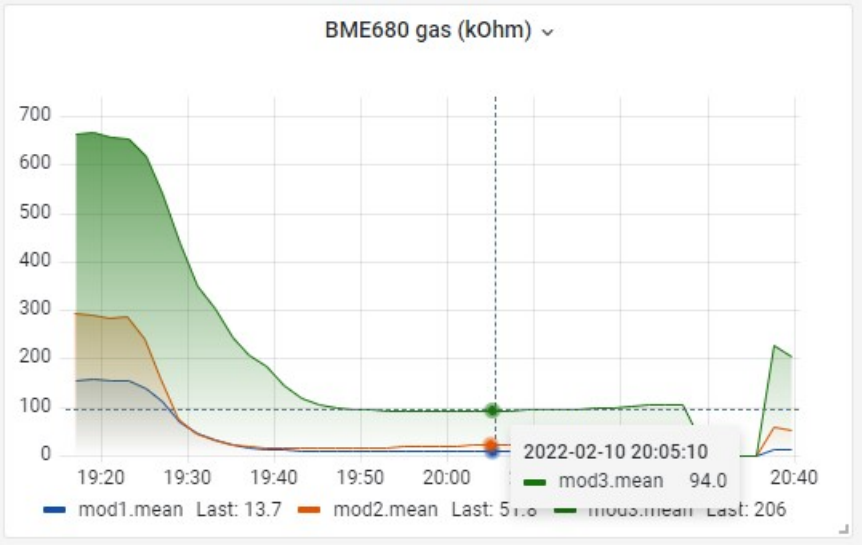

Figure 5.14: Resulting graph of BME680 gas level for smoke test.

Due to output signal behavior inversely proportional to gas level when starts the test BME680 sensor decrease this signal registered and keeps this behavior for all period.

5.4.2 Predicting test data

The algorithms were run using the test data, as expected, the prediction showed different values from the measured data. Figure 5.15 presents the MQ gas forecast, when analyzing the Error line, in both cases it is noted that the measured gas value has values greater than 1500 ppm than the predicted value. The assertively to LR and ANN models were respectively 46.61% and 34.1% and the mean absolute error scored around 629.42 ppm and 438.93 ppm.

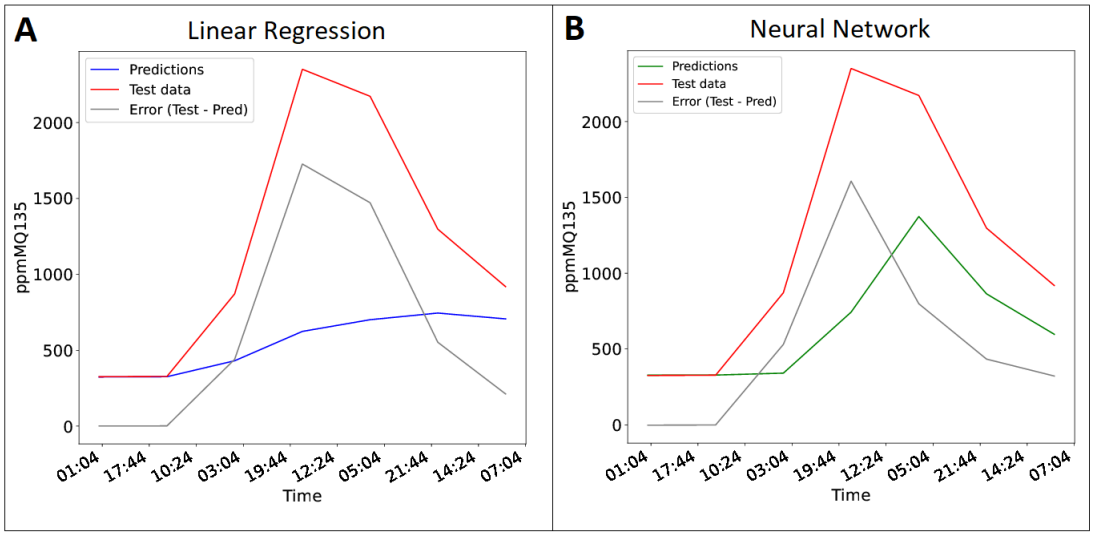

Figure 5.15: Prediction of MQ135 gas level for smoke test.

The same behavior happens with ZE08 gas predict, in Figure 5.16.A the Error line shows the difference between test and prediction values to the linear regression, the predict in this test was 47.24% and a median error from around 6.5 ppm. The prediction from ANN model in Figure 5.16.B generated only invalid values due to error and precision due to large difference between predicted data and measured data.

Analyzing the Figure 5.17, the forecast of BME680 gas concentration presents high differences between test and prediction values in both cases. The final prediction from LR and ANN scored respectively 60.63% and 61.2% and the average error was around 46.33 kΩ and 93.66 kΩ.

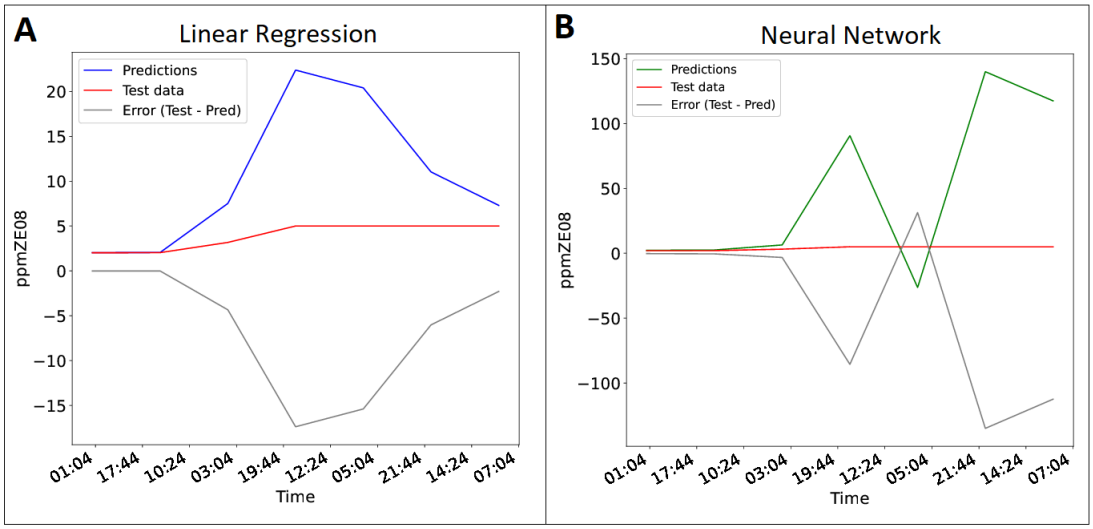

Figure 5.16: Prediction of ZE08 gas level for smoke test.

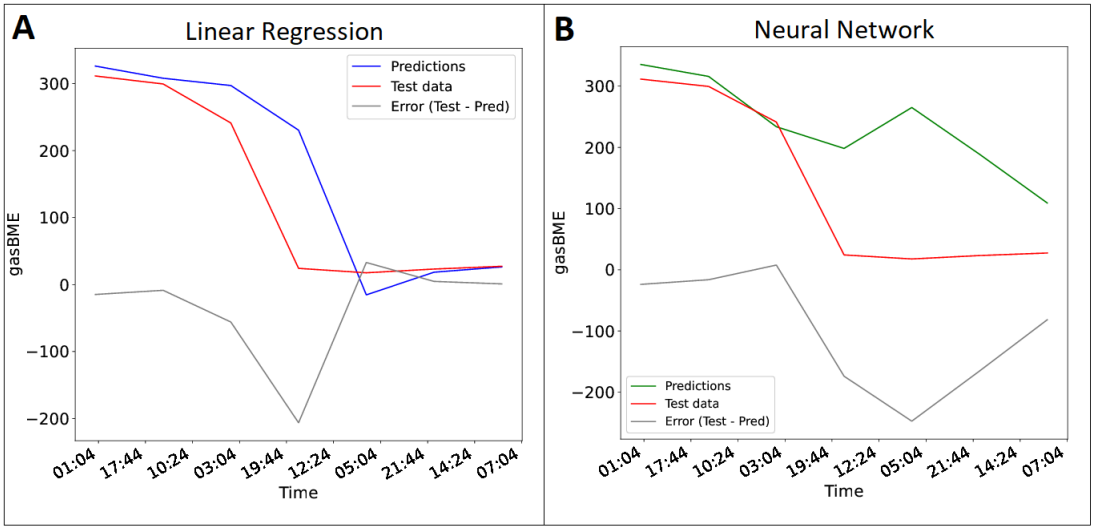

Figure 5.17: Prediction of BME680 gas level for smoke test.

In general, the smoke test presented expected results, given that the learning models were trained to predict the value of gases at normal levels according to the data stored in InfluxDB. This high value of detected error makes it possible to perceive situations that compromise the IAQ of the environment, thus being able to alert, in real time, about variations in the parameters.

5.5 Real-Time Monitoring

With the trained learning models, the real-time monitoring implementation is done in Grafana with a new database to store prediction values resulted from the LR and ANN models. Figure 5.18 is the real-time monitoring in normal conditions done in March 4th, the measured parameters are highlighted in blue are presenting a behavior similar to predictions values in yellow.

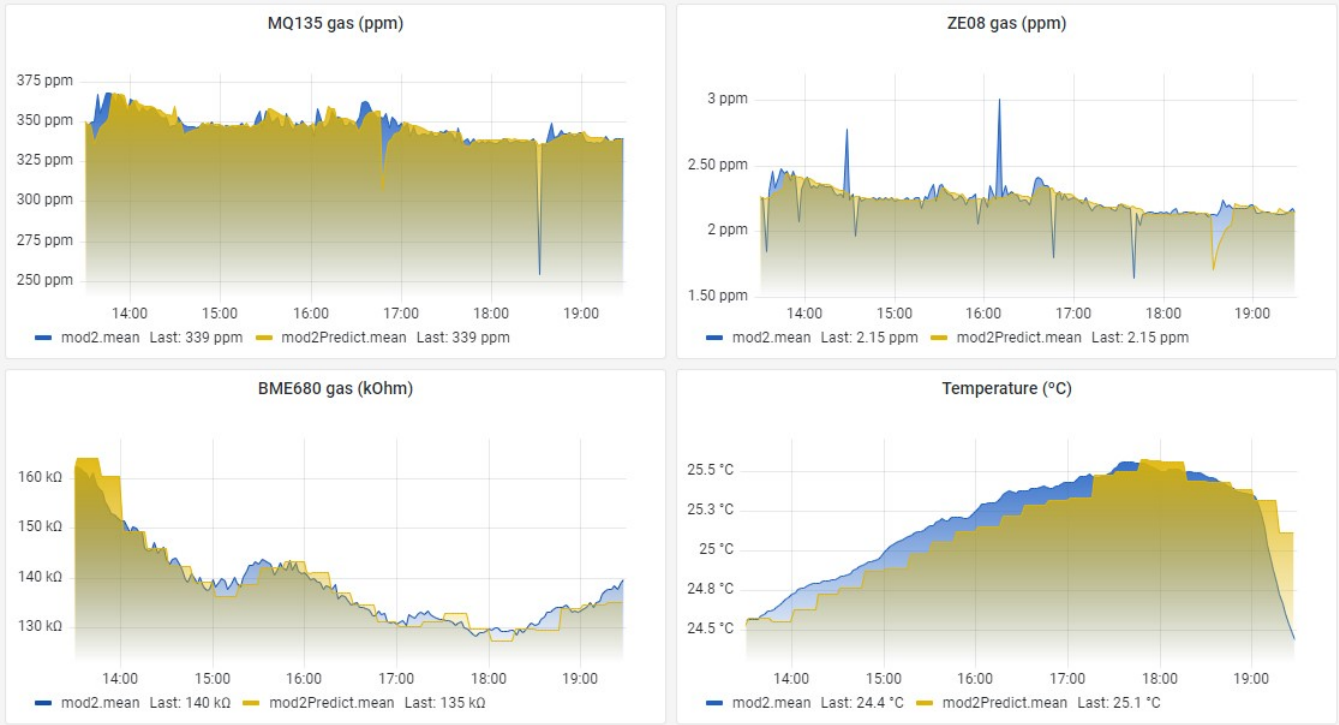

Figure 5.18: Real-time monitoring in normal conditions.

Figure 5.19 shows a monitoring of smoke test done to verify the real-time implementation where a dashboard with the high measured parameters on graphics, with the last measured values highlighted bellow, and the predictions of the next supposed value of each parameter displayed.

As explained in section 4.4, the data read by the sensors is sent to the database and automatically loaded into PyCharm IDE to be used in the learning algorithms, then the prediction values are sent to NodeRED and stored in a database. Then the data is displayed in Grafana in order to verify that the parameters read are as expected. The

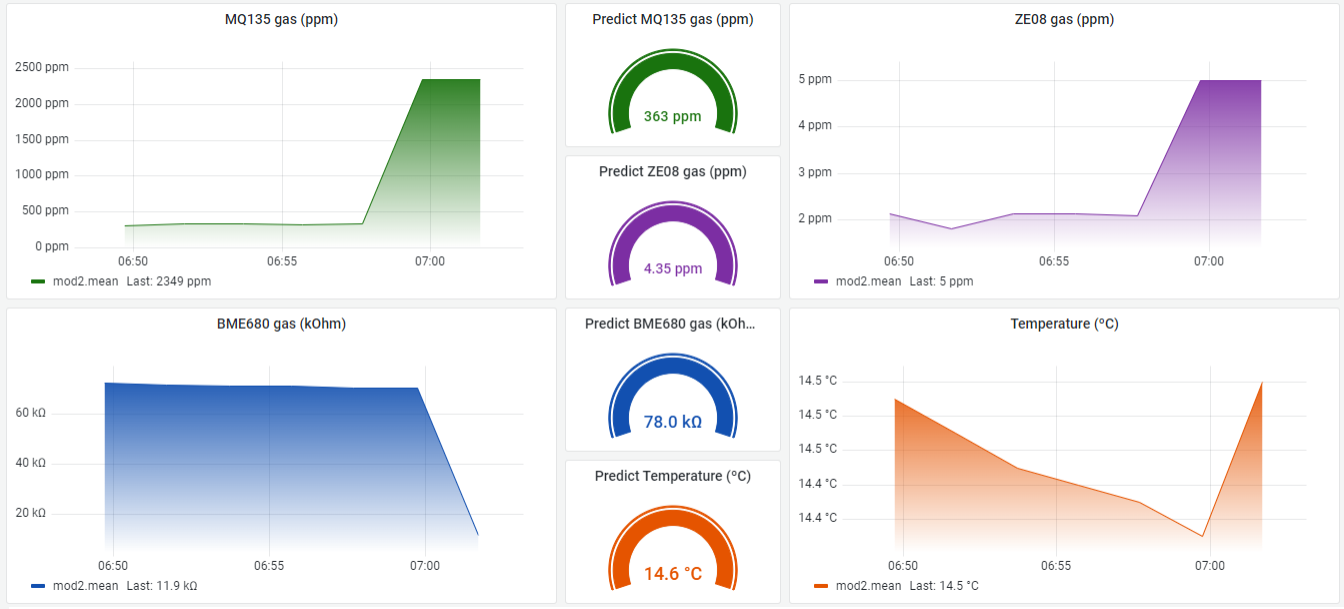

Figure 5.19: Dashboard to real-time monitoring.

entire process is repeated every two minutes to keep monitoring as up-to-date as possible for possible data variations.

Chapter 6

Conclusion

This work presents a real-time indoor air quality monitoring system, where two machine learning algorithms for data prediction were developed, being a mathematical linear regression model and an artificial neural network. The implementation process started with the choice of a set of low-cost sensors capable of measuring the main parameters responsible for the IAQ, such as CO, CO_2, VOCs, SO_2, O_3 and NO_x, in addition to temperature and humidity, which are comfort parameters that involve the IAQ. Then, the data were stored in an InfluxDB database, going through a process where they were first sent to NodeRED through the MQTT communication protocol and later forwarded to the database with the help of tools provided by NodeRED. The data were stored for a period of three months where they were constantly monitored through Grafana. This database was used to train forecasting models developed in PyCharm IDE and later used to compare against forecast results to assess the accuracy of learning models in forecasting. From the point of view of prediction efficiency, the linear regression model presents excellent results in well behaved situations, although in many studies the neural network model stands out in relation to the LR ([109] [110] [111] [115]), in the experiments performed, the mathematical model stood out in relation to the neural network, reaching values up to 99.7% accuracy when predicting the gas concentration measured by the BME680 sensor in usual situation. After the learning was done, they were tested in a smoke test to simulate a rising gas behavior in order to validate if the models would present errors

in the predictions, since they were only trained to predict the normal behavior of the measured parameters. When validating the functionality, the algorithms were adjusted to make predictions in real time and these values are displayed on a dashboard in Grafana with the actual values that are measured by the sensors to have a way to compare if the concentration of gases in the environment is following the behavior that should have given the predicted values.

6.1 Future Works

This work addresses several applications for the main topic that is IAQ monitoring. With this, some implementations for future work are presented:

- Include particle sensors in modules to improve the quality of data collected. The idea of using this type of sensor was still in mind for this work, however it was not possible due to supplier logistic problems;
- Improve machine learning models seeking more efficiency in the prediction of parameters and develop new algorithms in order to discover a model that best meets the requirements of the proposal;
- Develop a network of modules capable of communicating with each other and dividing them to study the behavior of parameters in different environments;
- Implement a control application for the system using the error signals between the measured values and the predicted values to trigger actuators such as ventilation and exhaust systems to stabilize ambient gas levels in the event of a leak;
- Development of an application to make the control and monitoring of the system more practical and accessible.

Bibliography

[1] J. Leech, W. C. Nelson, R. T. Burnett, S. D. Aaron, and M. E. Raizenne, "It's about time: A comparison of canadian and american time–activity patterns†," *Journal of Exposure Analysis and Environmental Epidemiology*, vol. 12, pp. 427–432, 2002.

[2] *WHO. Household air pollution and health*, https://www.who.int/en/news-room/fact-sheets/detail/household-air-pollution-and-health, Accessed: 2022-01-14.

[3] P. Kumar and B. Imam, "Footprints of air pollution and changing environment on the sustainability of built infrastructure," *Science of The Total Environment*, vol. 444, pp. 85–101, 2013, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2012.11.056. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969712014829.

[4] D. Ekmekcioglu and S. S. Keskin, "Characterization of indoor air particulate matter in selected elementary schools in istanbul, turkey," *Indoor and Built Environment*, vol. 16, no. 2, pp. 169–176, 2007. DOI: 10.1177/1420326X07076777. eprint: https://doi.org/10.1177/1420326X07076777. [Online]. Available: https://doi.org/10.1177/1420326X07076777.

[5] D. Pelufo and L. Durante, "Diretrizes para implantação de um sistema de saúde e segurança do trabalho em empresas produtoras de álcool em gel," *E&S - Engineering and Science*, vol. 1, no. 10, 2021, ISSN: 2358-5390. DOI: 10.18607/ES20211011776. [Online]. Available: https://periodicoscientificos.ufmt.br/ojs/index.php/eng/article/view/11776.

[6] M. L. Frenia and J. L. Schauben, "Methanol inhalation toxicity," *Annals of Emergency Medicine*, vol. 22, no. 12, pp. 1919–1923, 1993, ISSN: 0196-0644. DOI: https://doi.org/10.1016/S0196-0644(05)80424-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S019606440580424X.

[7] A. Persily, W. Dols, and S. Nabinger, "Air change effectiveness measurements in two modern office buildings," en, in Indoor Air, , -1, 1994-01-01 1994. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=916610.

[8] *Osha. technical manual: Indoor air quality investigation*, https://www.epa.gov/indoor-air-quality-iaq/indoorparticulate-matter, Accessed: 2022-01-14.

[9] S. Dhanalakshmi, M. Poongothai, and K. Sharma, "Iot based indoor air quality and smart energy management for hvac system," *Procedia Computer Science*, vol. 171, pp. 1800–1809, 2020, Third International Conference on Computing and Network Communications (CoCoNet'19), ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.04.193. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920311741.

[10] Z. Argunhan and A. Avci, "Statistical evaluation of indoor air quality parameters in classrooms of a university," *Advances in Meteorology*, vol. 2018, 2018. DOI: https://doi.org/10.1155/2018/4391579. [Online]. Available: https://www.hindawi.com/journals/amete/2018/4391579/#abstract.

[11] L. Fang, G. Clausen, and P. O. Fanger, "Impact of temperature and humidity on the perception of indoor air quality," *Indoor Air*, vol. 8, pp. 80–90, 1998.

[12] *VentPro. Indoor Air Quality The Pollution's*, https://www.ventpro.co.th/indoor-air-quality-the-pollutions/, Accessed: 2022-01-17.

[13] M. Marć, M. Śmiełowska, J. Namieśnik, and B. Zabiegała, "Indoor air quality of everyday use spaces dedicated to specific purposes—a review," *ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH*, vol. 25, no. 3, pp. 2065–2082,

2018. DOI: https://doi.org/10.1007/s11356-017-0839-8. [Online]. Available: https://link.springer.com/article/10.1007%5C%2Fs11356-017-0839-8#citeas.

[14] Z. Peng, W. Deng, and R. Tenorio, "Investigation of indoor air quality and the identification of influential factors at primary schools in the north of china," *Sustainability*, vol. 9, no. 7, 2017, ISSN: 2071-1050. DOI: 10.3390/su9071180. [Online]. Available: https://www.mdpi.com/2071-1050/9/7/1180.

[15] C. H. Linaker, A. J. Chauhan, H. Inskip, *et al.*, "Distribution and determinants of personal exposure to nitrogen dioxide in school children.," *Occupational and Environmental Medicine*, vol. 53, no. 3, pp. 200–203, 1996, ISSN: 1351-0711. DOI: 10.1136/oem.53.3.200. eprint: https://oem.bmj.com/content/53/3/200.full.pdf. [Online]. Available: https://oem.bmj.com/content/53/3/200.

[16] J. Ström, L. Alfredsson, T. Malmfors, and O. Selroos, "Review : Nitrogen dioxide: Causation and aggravation of lung diseases," *Indoor Environment*, vol. 3, no. 2, pp. 58–68, 1994. DOI: 10.1177/1420326X9400300202. eprint: https://doi.org/10.1177/1420326X9400300202. [Online]. Available: https://doi.org/10.1177/1420326X9400300202.

[17] M. Jafari, "Association of sick building syndrome with indoor air parameters," *Tanaffos*, vol. 14, pp. 55–62, Jan. 2015.

[18] P. Guo, K. Yokoyama, F. Piao, *et al.*, "Sick building syndrome by indoor air pollution in dalian, china," *International Journal of Environmental Research and Public Health*, vol. 10, no. 4, pp. 1489–1504, 2013, ISSN: 1660-4601. DOI: 10.3390/ijerph10041489. [Online]. Available: https://www.mdpi.com/1660-4601/10/4/1489.

[19] K. Kreiss, "The epidemiology of building-related complaints and illness.," *Occupational medicine*, vol. 4 4, pp. 575–92, 1989.

[20] S. Simkovich, D. Goodman, C. Roa, *et al.*, "The health and social implications of household air pollution and respiratory diseases.," *NPJ Prim. Care Respir. Med.*, vol. 29, no. 12, Apr. 2019. DOI: 10.1038/s41533-019-0126-x. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31028270/.

[21] S. N. Grief, "Upper respiratory infections," *Primary Care: Clinics in Office Practice*, vol. 40, no. 3, pp. 757–770, 2013, Infectious Disease, ISSN: 0095-4543. DOI: https://doi.org/10.1016/j.pop.2013.06.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095454313000663.

[22] H. Jary, H. Simpson, D. Havens, *et al.*, "Household air pollution and acute lower respiratory infections in adults: A systematic review," *PLOS ONE*, vol. 11, no. 12, pp. 1–14, Dec. 2016. DOI: 10.1371/journal.pone.0167656. [Online]. Available: https://doi.org/10.1371/journal.pone.0167656.

[23] O. Raaschou-Nielsen, R. Beelen, M. Wang, *et al.*, "Particulate matter air pollution components and risk for lung cancer," *Environment International*, vol. 87, pp. 66–73, 2016, ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2015.11.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016041201530091X.

[24] J. Kirkby, V. Bountziouka, S. Lum, A. Wade, and J. Stocks, "Natural variability of lung function in young healthy school children," *European Respiratory Journal*, vol. 48, no. 2, pp. 411–419, 2016, ISSN: 0903-1936. DOI: 10.1183/13993003.01795-2015. eprint: https://erj.ersjournals.com/content/48/2/411.full.pdf. [Online]. Available: https://erj.ersjournals.com/content/48/2/411.

[25] K.-H. Kim, S. A. Jahan, and E. Kabir, "A review of diseases associated with household air pollution due to the use of biomass fuels," *Journal of Hazardous Materials*, vol. 192, no. 2, pp. 425–431, 2011, ISSN: 0304-3894. DOI: https://doi.org/10.1016/j.jhazmat.2011.05.087. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304389411007424.

[26] M.-i. Choi, K. Cho, J. Y. Hwang, *et al.*, "Design and implementation of iot-based hvac system for future zero energy building," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017, pp. 605–610. DOI: 10.1109/PERCOMW.2017.7917631.

[27] J. M. Samet, M. C. Marbury, and J. D. Spengler, "Health effects and sources of indoor air pollution. part i," *The American review of respiratory disease*, vol. 6, no. 136, pp. 1919–1923, 1987, ISSN: 1486–1508. DOI: 10.1164/ajrccm/136.6.1486. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3318602/.

[28] C. S. Mitchell, J. (Zhang, T. Sigsgaard, *et al.*, "Current state of the science: Health effects and indoor environmental quality," *Environmental Health Perspectives*, vol. 115, no. 6, pp. 958–964, 2007. DOI: 10.1289/ehp.8987. eprint: https://ehp.niehs.nih.gov/doi/pdf/10.1289/ehp.8987. [Online]. Available: https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.8987.

[29] V. V. Tran, D. Park, and Y.-C. Lee, "Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, 2020, ISSN: 1660-4601. DOI: 10.3390/ijerph17082927. [Online]. Available: https://www.mdpi.com/1660-4601/17/8/2927.

[30] L. S. R. Brickus, J. N. Cardoso, and F. R. de Aquino Neto, "Distributions of indoor and outdoor air pollutants in rio de janeiro, brazil: Implications to indoor air quality in bayside offices," *Environmental Science & Technology*, vol. 32, no. 22, pp. 3485–3490, 1998. DOI: 10.1021/es980336x. eprint: https://doi.org/10.1021/es980336x. [Online]. Available: https://doi.org/10.1021/es980336x.

[31] C. Lombardi, C. Spedini, and S. Govoni, "Effect of calcium entry blockade on ethanol-induced changes in bronchomotor tone," *European Journal of Clinical Pharmacology*, vol. 28, pp. 221–222, 2004.

[32] M. R. Kinshella, M. V. V. Dyke, K. E. Douglas, and J. W. Martyny, "Perceptions of indoor air quality associated with ventilation system types in elementary schools," *Applied Occupational and Environmental Hygiene*, vol. 16, no. 10, pp. 952–960, 2001, PMID: 11599544. DOI: 10.1080/104732201300367209. eprint: https://doi.org/10.1080/104732201300367209. [Online]. Available: https://doi.org/10.1080/104732201300367209.

[33] T. Putus, A. Tuomainen, and S. Rautiala, "Chemical and microbial exposures in a school building: Adverse health effects in children," *Archives of Environmental Health: An International Journal*, vol. 59, no. 4, pp. 194–201, 2004, PMID: 16189992. DOI: 10.3200/AEOH.59.4.194-201. eprint: https://doi.org/10.3200/AEOH.59.4.194-201. [Online]. Available: https://doi.org/10.3200/AEOH.59.4.194-201.

[34] Y.-H. Mi, D. Norbäck, J. Tao, Y.-L. Mi, and M. Ferm, "Current asthma and respiratory symptoms among pupils in shanghai, china: Influence of building ventilation, nitrogen dioxide, ozone, and formaldehyde in classrooms," *Indoor air*, vol. 16, no. 6, pp. 454–464, Dec. 2006, ISSN: 0905-6947. DOI: 10.1111/j.1600-0668.2006.00439.x. [Online]. Available: https://doi.org/10.1111/j.1600-0668.2006.00439.x.

[35] Z. Zhao, A. Sebastian, L. Larsson, Z. Wang, Z. Zhang, and D. Norbäck, "Asthmatic symptoms among pupils in relation to microbial dust exposure in schools in taiyuan, china," *Pediatric Allergy and Immunology*, vol. 19, no. 5, pp. 455–465, 2008.

[36] H. Altuğ, E. O. Gaga, T. Döğeroğlu, *et al.*, "Effects of air pollution on lung function and symptoms of asthma, rhinitis and eczema in primary school children," *Environmental Science and Pollution Research*, vol. 20, no. 9, pp. 6455–6467, 2013.

[37] M. Turunen, O. Toyinbo, T. Putus, A. Nevalainen, R. Shaughnessy, and U. Haverinen-Shaughnessy, "Indoor environmental quality in school buildings, and the health

and wellbeing of students," *International journal of hygiene and environmental health*, vol. 217, no. 7, pp. 733–739, 2014.

[38] U. Haverinen-Shaughnessy, R. J. Shaughnessy, E. C. Cole, O. Toyinbo, and D. J. Moschandreas, "An assessment of indoor environmental quality in schools and its association with health and performance," *Building and Environment*, vol. 93, pp. 35–40, 2015.

[39] V. Chithra and S. Shiva Nagendra, "A review of scientific evidence on indoor air of school building: Pollutants, sources, health effects and management," *Asian J. Atmos. Environ*, vol. 12, no. 2, 2018. DOI: https://doi.org/10.5572/ajae.2018.12.2.87. [Online]. Available: https://www.koreascience.or.kr/article/JAKO201821142176514.page.

[40] X.-Q. Jiang, X.-D. Mei, and D. Feng, "Air pollution and chronic airway diseases: What should people know and do?" *Journal of Thoracic Disease*, vol. 8, no. 1, 2015, ISSN: 2077-6624. [Online]. Available: https://jtd.amegroups.com/article/view/5951.

[41] H. Qiu, I. T.-s. Yu, L. Tian, *et al.*, "Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: A time-series analysis in hong kong," *Environmental Health Perspectives*, vol. 120, no. 4, pp. 572–576, 2012. DOI: 10.1289/ehp.1104002. eprint: https://ehp.niehs.nih.gov/doi/pdf/10.1289/ehp.1104002. [Online]. Available: https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.1104002.

[42] F. W. S. Ko, W. Tam, T. W. Wong, *et al.*, "Effects of air pollution on asthma hospitalization rates in different age groups in hong kong," *Clinical & Experimental Allergy*, vol. 37, no. 9, pp. 1312–1319, 2007. DOI: https://doi.org/10.1111/j.1365-2222.2007.02791.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2222.2007.02791.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2222.2007.02791.x.

[43] M.-H. Cheng, C.-C. Chen, H.-F. Chiu, and C.-Y. Yang, "Fine particulate air pollution and hospital admissions for asthma: A case-crossover study in taipei," *Journal of Toxicology and Environmental Health, Part A*, vol. 77, no. 18, pp. 1075–1083, 2014, PMID: 25072894. DOI: 10.1080/15287394.2014.922387. eprint: https://doi.org/10.1080/15287394.2014.922387. [Online]. Available: https://doi.org/10.1080/15287394.2014.922387.

[44] Y. Zhao, S. Wang, K. Aunan, H. Martin Seip, and J. Hao, "Air pollution and lung cancer risks in china—a meta-analysis," *Science of The Total Environment*, vol. 366, no. 2, pp. 500–513, 2006, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2005.10.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969705007679.

[45] Y. Guo, H. Zeng, R. Zheng, *et al.*, "The burden of lung cancer mortality attributable to fine particles in china," *Science of The Total Environment*, vol. 579, pp. 1460–1466, 2017, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2016.11.147. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969716326018.

[46] M. S. Burroughs Peña and A. Rollins, "Environmental exposures and cardiovascular disease: A challenge for health and development in low- and middle-income countries," *Cardiology Clinics*, vol. 35, no. 1, pp. 71–86, 2017, Global Cardiovascular Health, ISSN: 0733-8651. DOI: https://doi.org/10.1016/j.ccl.2016.09.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0733865116300765.

[47] M. Ezzati, "Indoor air pollution and health in developing countries," *The Lancet*, vol. 366, no. 9480, pp. 104–106, 2005, ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(05)66845-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673605668456.

[48] C. H. Backes, T. Nelin, M. W. Gorr, and L. E. Wold, "Early life exposure to air pollution: How bad is it?" *Toxicology Letters*, vol. 216, no. 1, pp. 47–53, 2013,

ISSN: 0378-4274. DOI: https://doi.org/10.1016/j.toxlet.2012.11.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037842741201380X.

[49] V. Mishra, X. Dai, K. R. Smith, and L. Mika, "Maternal exposure to biomass smoke and reduced birth weight in zimbabwe," *Annals of Epidemiology*, vol. 14, no. 10, pp. 740–747, 2004, ISSN: 1047-2797. DOI: https://doi.org/10.1016/j.annepidem.2004.01.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047279704000456.

[50] P. Lakshmi, N. K. Virdi, A. Sharma, *et al.*, "Household air pollution and stillbirths in india: Analysis of the dlhs-ii national survey," *Environmental Research*, vol. 121, pp. 17–22, 2013, ISSN: 0013-9351. DOI: https://doi.org/10.1016/j.envres.2012.12.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935112003246.

[51] A. K. Amegah, S. Näyhä, and J. J. K. Jaakkola, "Do biomass fuel use and consumption of unsafe water mediate educational inequalities in stillbirth risk? an analysis of the 2007 ghana maternal health survey," *BMJ Open*, vol. 7, no. 2, 2017, ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-012348. eprint: https://bmjopen.bmj.com/content/7/2/e012348.full.pdf. [Online]. Available: https://bmjopen.bmj.com/content/7/2/e012348.

[52] A. Kankaria, B. Nongkynrih, and S. Gupta, "Indoor air pollution in India: Implications on health and its control," *Indian Journal of Community Medicine*, vol. 39, no. 4, pp. 203–207, 2014. DOI: 10.4103/0970-0218.143019. eprint: https://www.ijcm.org.in/article.asp?issn=0970-0218;year=2014;volume=39;issue=4;spage=203;epage=207;aulast=Kankaria;t=6. [Online]. Available: https://www.ijcm.org.in/article.asp?issn=0970-0218;year=2014;volume=39;issue=4;spage=203;epage=207;aulast=Kankaria;t=6.

[53] D. P. Pope, V. Mishra, L. Thompson, *et al.*, "Risk of Low Birth Weight and Stillbirth Associated With Indoor Air Pollution From Solid Fuel Use in Developing

Countries," *Epidemiologic Reviews*, vol. 32, no. 1, pp. 70–81, Apr. 2010, ISSN: 0193-936X. DOI: 10.1093/epirev/mxq005. eprint: https://academic.oup.com/epirev/article-pdf/32/1/70/1339977/mxq005.pdf. [Online]. Available: https://doi.org/10.1093/epirev/mxq005.

[54] M. Ezzati, J. Utzinger, S. Cairncross, A. J. Cohen, and B. H. Singer, "Environmental risks in the developing world: Exposure indicators for evaluating interventions, programmes, and policies," *Journal of Epidemiology & Community Health*, vol. 59, no. 1, pp. 15–22, 2005, ISSN: 0143-005X. DOI: 10.1136/jech.2003.019471. eprint: https://jech.bmj.com/content/59/1/15.full.pdf. [Online]. Available: https://jech.bmj.com/content/59/1/15.

[55] G. A. Stevens, R. H. Dias, and M. Ezzati, "The effects of 3 environmental risks on mortality disparities across mexican communities," *Proceedings of the National Academy of Sciences*, vol. 105, no. 44, pp. 16 860–16 865, 2008, ISSN: 0027-8424. DOI: 10.1073/pnas.0808927105. eprint: https://www.pnas.org/content/105/44/16860.full.pdf. [Online]. Available: https://www.pnas.org/content/105/44/16860.

[56] J. Li, S.-W. Yin, G.-S. Shi, and L. Wang, "Optimization of indoor thermal comfort parameters with the adaptive network-based fuzzy inference system and particle swarm optimization algorithm," *Mathematical Problems in Engineering*, vol. 2017, 2017. DOI: https://doi.org/10.1155/2017/3075432. [Online]. Available: https://www.hindawi.com/journals/mpe/2017/3075432/#abstract.

[57] J. Saini, M. Dutta, and G. Marques, "Indoor air quality monitoring systems based on internet of things: A systematic review," *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, 2020, ISSN: 1660-4601. DOI: 10.3390/ijerph17144942. [Online]. Available: https://www.mdpi.com/1660-4601/17/14/4942.

[58] M. Braik, A. Sheta, and H. Al-Hiary, "Hybrid neural network models for forecasting ozone and particulate matter concentrations in the republic of china," *Air*

Qual Atmos Health, vol. 13, 2020, ISSN: 839–851. DOI: https://doi.org/10.1007/s11869-020-00841-7. [Online]. Available: https://link.springer.com/article/10.1007%5C%2Fs11869-020-00841-7#citeas.

[59] F. X. Ming, R. A. A. Habeeb, F. H. B. Md Nasaruddin, and A. B. Gani, "Real-time carbon dioxide monitoring based on iot & cloud technologies," in *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, Penang, Malaysia: Association for Computing Machinery, 2019, pp. 517–521, ISBN: 9781450365734. DOI: 10.1145/3316615.3316622. [Online]. Available: https://doi.org/10.1145/3316615.3316622.

[60] X. Yang, Q. Chen, J. Zhang, Y. An, J. Zeng, and C. Shaw, "A mass transfer model for simulating voc sorption on building materials," *Atmospheric Environment*, vol. 35, no. 7, pp. 1291–1299, 2001, ISSN: 1352-2310. DOI: https://doi.org/10.1016/S1352-2310(00)00397-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231000003976.

[61] J. Saini, M. Dutta, and G. Marques, "Sensors for indoor air quality monitoring and assessment through internet of things: A systematic review," *Environmental Monitoring and Assessment*, vol. 193, no. 2, pp. 1–32, 2021.

[62] E. Alexandrova and A. Ahmadinia, "Real-time intelligent air quality evaluation on a resource-constrained embedded platform," in *2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS)*, 2018, pp. 165–170. DOI: 10.1109/BDS/HPSC/IDS18.2018.00045.

[63] M. Muladi, S. Sendari, and T. Widiyaningtyas, "Real time indoor air quality monitoring using internet of things at university," in *2018 2nd Borneo International Conference on Applied Mathematics and Engineering (BICAME)*, 2018, pp. 169–173. DOI: 10.1109/BICAME45512.2018.1570509614.

[64] S. C. Folea and G. D. Mois, "Lessons learned from the development of wireless environmental sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3470–3480, 2020. DOI: 10.1109/TIM.2019.2938137.

[65] G. Chiesa, S. Cesari, M. Garcia, M. Issa, and S. Li, "Multisensor iot platform for optimising iaq levels in buildings through a smart ventilation system," *Sustainability*, vol. 11, no. 20, 2019, ISSN: 2071-1050. DOI: 10.3390/su11205777. [Online]. Available: https://www.mdpi.com/2071-1050/11/20/5777.

[66] Y. Tong, X. Mei-de, Y. Zi-han, and Z. Tian-qing, "Design of air quality monitoring system based on light scattering sensor," *IOP Conference Series: Earth and Environmental Science*, vol. 647, no. 1, p. 012 196, Jan. 2021. DOI: 10.1088/1755-1315/647/1/012196. [Online]. Available: https://doi.org/10.1088/1755-1315/647/1/012196.

[67] L. Russi, P. Guidorzi, B. Pulvirenti, G. Semprini, D. Aguiari, and G. Pau, "Air quality and comfort characterisation within an electric vehicle cabin," in *2021 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, 2021, pp. 169–174. DOI: 10.1109/MetroAutomotive50197.2021.9502853.

[68] X. Yang, L. Yang, and J. Zhang, "A wifi-enabled indoor air quality monitoring and control system: The design and control experiments," in *2017 13th IEEE International Conference on Control Automation (ICCA)*, 2017, pp. 927–932. DOI: 10.1109/ICCA.2017.8003185.

[69] M. Firdhous, B. Sudantha, and P. Karunaratne, "Iot enabled proactive indoor air quality monitoring system for sustainable health management," in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, 2017, pp. 216–221. DOI: 10.1109/ICCCT2.2017.7972281.

[70] Q. Yang, G. Zhou, W. Qin, B. Zhang, and P. Y. Chiang, "Air-kare: A wi-fi based, multi-sensor, real-time indoor air quality monitor," in *2015 IEEE International Wireless Symposium (IWS 2015)*, 2015, pp. 1–4. DOI: 10.1109/IEEE-IWS.2015.7164542.

[71] N. Naik, "Choice of effective messaging protocols for iot systems: Mqtt, coap, amqp and http," in *2017 IEEE International Systems Engineering Symposium (ISSE)*, 2017, pp. 1–7. DOI: 10.1109/SysEng.2017.8088251.

[72] J. Saini, M. Dutta, and G. Marques, "Indoor air quality monitoring systems and covid-19," in *Emerging Technologies During the Era of COVID-19 Pandemic*, Springer, 2021, pp. 133–147.

[73] P. Spachos and D. Hatzinakos, "Real-time indoor carbon dioxide monitoring through cognitive wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 2, pp. 506–514, 2016. DOI: 10.1109/JSEN.2015.2479647.

[74] G. Marques, C. R. Ferreira, and R. Pitarma, "Indoor air quality assessment using a co2 monitoring system based on internet of things," *Journal of medical systems*, vol. 43, no. 3, pp. 1–10, 2019.

[75] M. Firdhous, B. Sudantha, and P. Karunaratne, "Iot enabled proactive indoor air quality monitoring system for sustainable health management," in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, IEEE, 2017, pp. 216–221.

[76] M. Benammar, A. Abdaoui, S. H. Ahmad, F. Touati, and A. Kadri, "A modular iot platform for real-time indoor air quality monitoring," *Sensors*, vol. 18, no. 2, p. 581, 2018.

[77] N. Q. Pham, V. P. Rachim, and W.-Y. Chung, "Emi-free bidirectional real-time indoor environment monitoring system," *IEEE Access*, vol. 7, pp. 5714–5722, 2018.

[78] A. Chamseddine, I. Alameddine, M. Hatzopoulou, and M. El-Fadel, "Seasonal variation of air quality in hospitals with indoor–outdoor correlations," *Building and Environment*, vol. 148, pp. 689–700, 2019.

[79] Z. Liu, G. Wang, L. Zhao, and G. Yang, "Multi-points indoor air quality monitoring based on internet of things," *IEEE Access*, vol. 9, pp. 70 479–70 492, 2021. DOI: 10.1109/ACCESS.2021.3073681.

[80] R. M. Nica, T. Hapurne, A. I. Dumitrascu, I. Bliuc, and C. Avram, "Proposal for a small two-story living room house based on air-quality monitoring," *Procedia Manufacturing*, vol. 22, pp. 268–273, 2018.

[81] A. A. Hapsari, A. I. Hajamydeen, D. J. Vresdian, M. Manfaluthy, L. Prameswono, and E. Yusuf, "Real time indoor air quality monitoring system based on iot using mqtt and wireless sensor network," in *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 2019, pp. 1–7. DOI: 10.1109/ICETAS48360.2019.9117518.

[82] P. Asthana and S. Mishra, "Iot enabled real time bolt based indoor air quality monitoring system," in *2018 International Conference on Computational and Characterization Techniques in Engineering Sciences (CCTES)*, 2018, pp. 36–39. DOI: 10.1109/CCTES.2018.8674076.

[83] A. A. Hapsari, D. Junesco Vresdian, M. Aldiansyah, B. W. Dionova, and A. C. Windari, "Indoor air quality monitoring system with node.js and mqtt application," in *2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE)*, 2020, pp. 144–149. DOI: 10.1109/ICITAMEE50454.2020.9398324.

[84] L. Russi, P. Guidorzi, B. Pulvirenti, D. Aguiari, G. Pau, and G. Semprini, "Air quality and comfort characterisation within an electric vehicle cabin in heating and cooling operations," *Sensors*, vol. 22, no. 2, 2022, ISSN: 1424-8220. DOI: 10.3390/s22020543. [Online]. Available: https://www.mdpi.com/1424-8220/22/2/543.

[85] P. Lasomsri, P. Yanbuaban, O. Kerdpoca, and T. Ouypornkochagorn, "A development of low-cost devices for monitoring indoor air quality in a large-scale hospital," in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2018, pp. 282–285. DOI: 10.1109/ECTICon.2018.8619934.

[86] A. D. I. A. Kadir, M. R. N. M. Alias, D. R. M. Dzaki, N. M. Din, S. N. M. Deros, and M. H. Haron, "Cloud-based iot air quality monitoring system," in *2021 26th IEEE Asia-Pacific Conference on Communications (APCC)*, 2021, pp. 121–127. DOI: 10.1109/APCC49754.2021.9609897.

[87] T. H. Nasution, A. Hizriadi, K. Tanjung, and F. Nurmayadi, "Design of indoor air quality monitoring systems," in *2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, IEEE, 2020, pp. 238–241.

[88] B. K. Moharana, P. Anand, S. Kumar, and P. Kodali, "Development of an iot-based real-time air quality monitoring device," in *2020 International conference on communication and signal processing (ICCSP)*, IEEE, 2020, pp. 191–194.

[89] S. McGrath, C. Flanagan, L. Zeng, and C. O'Leary, "Iot personal air quality monitor," in *2020 31st Irish signals and systems conference (ISSC)*, IEEE, 2020, pp. 1–4.

[90] P. G. Agbulu and G. J. R. Kumar, "An ultra-low power IoT system for indoor air quality monitoring," *Journal of Physics: Conference Series*, vol. 2007, no. 1, p. 012 053, Aug. 2021. DOI: 10.1088/1742-6596/2007/1/012053. [Online]. Available: https://doi.org/10.1088/1742-6596/2007/1/012053.

[91] M. Jacob Rodrigues, O. Postolache, and F. Cercas, "Indoor air quality monitoring system to prevent the triggering of respiratory distress," in *2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI)*, 2019, pp. 1–6. DOI: 10.1109/ISSI47111.2019.9043669.

[92] F. Pradityo and N. Surantha, "Indoor air quality monitoring and controlling system based on iot and fuzzy logic," in *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 2019, pp. 1–6. DOI: 10.1109/ICoICT.2019.8835246.

[93] M. Mossolly, K. Ghali, and N. Ghaddar, "Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm," *Energy*, vol. 34, no. 1, pp. 58–66, 2009.

[94] M. Bagheri, A. Akbari, and S. A. Mirbagheri, "Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review," *Process Safety and Environmental Protection*, vol. 123, pp. 229–252, 2019.

[95] M. Wernick, Y. Yang, J. Brankov, G. Yourganov, and S. Strother, "Drawing conclusions from medical images," *IEEE Signal Process. Mag.*, vol. 25, pp. 25–38, 2010.

[96] T. W. Ayele and R. Mehta, "Air pollution monitoring and prediction using iot," in *2018 second international conference on inventive communication and computational technologies (ICICCT)*, IEEE, 2018, pp. 1741–1745.

[97] J. Saini, M. Dutta, and G. Marques, "Indoor air quality monitoring with iot: Predicting pm10 for enhanced decision support," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 2020, pp. 504–508. DOI: 10.1109/DASA51403.2020.9317054.

[98] ——, "Internet of things based environment monitoring and pm<inf>10</inf> prediction for smart home," in *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, 2020, pp. 1–5. DOI: 10.1109/3ICT51146.2020.9311996.

[99] P. R. Meris, E. Dimaunahan, J. C. Dela Cruz, *et al.*, "Iot based – automated indoor air quality and lpg leak detection control system using support vector machine," in *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, 2020, pp. 231–235. DOI: 10.1109/ICSGRC49013.2020.9232472.

[100] J. Li, J. Wall, and G. Platt, "Indoor air quality control of hvac system," in *Proceedings of the 2010 International Conference on Modelling, Identification and Control*, 2010, pp. 756–761.

[101] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin, "Reducing transient and steady state electricity consumption in hvac using learning-based model-predictive control," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 240–253, 2011.

[102] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models," in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 3438–3448. DOI: 10.1109/WSC.2015.7408504.

[103] S. Ghosal, S. Sengupta, M. Majumder, and B. Sinha, "Linear regression analysis to predict the number of deaths in india due to sars-cov-2 at 6 weeks from day 0 (100 cases-march 14th 2020)," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 311–315, 2020.

[104] T. Catalina, J. Virgone, and E. Blanco, "Development and validation of regression models to predict monthly heating demand for residential buildings," *Energy and buildings*, vol. 40, no. 10, pp. 1825–1832, 2008.

[105] A. E. Ben-Nakhi and M. A. Mahmoud, "Cooling load prediction for buildings using general regression neural networks," *Energy Conversion and Management*, vol. 45, no. 13-14, pp. 2127–2141, 2004.

[106] B. B. Ekici and U. T. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 356–362, 2009.

[107] R. Mena, F. Rodriguez, M. Castilla, and M. R. Arahal, "A prediction model based on neural networks for the energy consumption of a bioclimatic building," *Energy and Buildings*, vol. 82, pp. 142–155, 2014.

[108] R. Platon, V. R. Dehkordi, and J. Martel, "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," *Energy and Buildings*, vol. 92, pp. 10–18, 2015.

[109] S. Farzana, M. Liu, A. Baldwin, and M. U. Hossain, "Multi-model prediction and simulation of residential building energy in urban areas of chongqing, south west china," *Energy and Buildings*, vol. 81, pp. 161–169, 2014.

[110] Y. Zhang, Z. O'Neill, B. Dong, and G. Augenbroe, "Comparisons of inverse modeling approaches for predicting building energy performance," *Building and Environment*, vol. 86, pp. 177–190, 2015.

[111] H. Xie, F. Ma, and Q. Bai, "Prediction of indoor air quality using artificial neural networks," in *2009 Fifth International Conference on Natural Computation*, vol. 2, 2009, pp. 414–418. DOI: 10.1109/ICNC.2009.502.

[112] C. Ren and S.-J. Cao, "Implementation and visualization of artificial intelligent ventilation control system using fast prediction models and limited monitoring data," *Sustainable Cities and Society*, vol. 52, p. 101 860, 2020, ISSN: 2210-6707. DOI: https://doi.org/10.1016/j.scs.2019.101860. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670719325302.

[113] R. Mumtaz, S. M. H. Zaidi, M. Z. Shakir, *et al.*, "Internet of things (iot) based indoor air quality sensing and predictive analytic—a covid-19 perspective," *Electronics*, vol. 10, no. 2, 2021, ISSN: 2079-9292. DOI: 10.3390/electronics10020184. [Online]. Available: https://www.mdpi.com/2079-9292/10/2/184.

[114] X. Chen, Y. Zheng, Y. Chen, *et al.*, "Indoor air quality monitoring system for smart buildings," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14, Seattle, Washington: Association for Computing Machinery, 2014, pp. 471–475, ISBN: 9781450329682. DOI: 10.1145/2632048.2632103. [Online]. Available: https://doi.org/10.1145/2632048.2632103.

[115] H. Zhang, R. Srinivasan, and X. Yang, "Simulation and analysis of indoor air quality in florida using time series regression (tsr) and artificial neural networks (ann) models," *Symmetry*, vol. 13, no. 6, 2021, ISSN: 2073-8994. DOI: 10.3390/sym13060952. [Online]. Available: https://www.mdpi.com/2073-8994/13/6/952.

[116] E. System, "Esp32 series datasheet," Tech. Rep., Aug. 2021.

[117] *MOUSER ELECTRONICS - ESP32*, https://pt.mouser.com/ProductDetail/Espressif-Systems/ESP32-DevKitC-32UE?qs=GedFDFLaBXFguOYDKoZ3jA%3D%3D, Accessed: 2022-02-15.

[118] E. Systems, "Esp32-wroom-32 datasheet," Tech. Rep., Aug. 2021.

[119] B. Sensortec, "Bme680 low power gas, pressure, temperature & humidity sensor," Tech. Rep., Jul. 2017.

[120] *SATKIT - BME680 sensor*, https://satkit.com/pt-pt/bme680-sensor-de-gs-de-presso-de-ar-e-umidade-i2c-arduino-raspberry-pi, Accessed: 2022-02-15.

[121] Winsen, "Mq135 semiconductor sensor for air quality," Tech. Rep., Mar. 2015.

[122] W. sensor, "Electrochemical ch2o detection module," Tech. Rep., Dez 2014.

[123] *Arduino*, https://www.arduino.cc/, Accessed: 2021-08-17.

[124] N. Naik, "Choice of effective messaging protocols for iot systems: Mqtt, coap, amqp and http," in *2017 IEEE International Systems Engineering Symposium (ISSE)*, 2017, pp. 1–7. DOI: 10.1109/SysEng.2017.8088251.

[125] N. Havard, S. McGrath, C. Flanagan, and C. MacNamee, "Smart building based on internet of things technology," in *2018 12th International Conference on Sensing Technology (ICST)*, 2018, pp. 278–281. DOI: 10.1109/ICSensT.2018.8603575.

[126] *Influxdata*, https://www.influxdata.com/, Accessed: 2021-08-17.

[127] *Grafana Labs*, https://grafana.com/, Accessed: 2021-08-17.

[128] W. Wei, O. Ramalho, L. Malingre, S. Sivanantham, J. C. Little, and C. Mandin, "Machine learning and statistical models for predicting indoor air quality," *Indoor Air*, vol. 29, no. 5, pp. 704–726, 2019.

[129] B. Khazaei, A. Shiehbeigi, and A. Haji Molla Ali Kani, "Modeling indoor air carbon dioxide concentration using artificial neural network," *International journal of environmental science and technology*, vol. 16, no. 2, pp. 729–736, 2019.

[130] A. Challoner, F. Pilla, and L. Gill, "Prediction of indoor air exposure from outdoor air quality using an artificial neural network model for inner city commercial buildings," *International journal of environmental research and public health*, vol. 12, no. 12, pp. 15 233–15 253, 2015.

[131] *PyCharm*, `https://www.jetbrains.com/pycharm/`, Accessed: 2022-02-04.

[132] A. S. Ali, C. Coté, M. Heidarinejad, and B. Stephens, "Elemental: An open-source wireless hardware and software platform for building energy and indoor environmental monitoring and control," *Sensors*, vol. 19, no. 18, p. 4017, 2019.

[133] A. Y. Sun, Z. Zhong, H. Jeong, and Q. Yang, "Building complex event processing capability for intelligent environmental monitoring," *Environmental modelling & software*, vol. 116, pp. 1–6, 2019.

[134] H. Nigam, A. K. Saini, S. Banerjee, and A. Kumar, "Indoor environment air quality monitoring and its notification to building occupants," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 2444–2448. DOI: `10.1109/TENCON.2019.8929232`.

Appendix A

Appendices

In this section, the pseudocode of the linear regression algorithm is represented in Figures A.1 and A.2. Then, in Figure A.3 the complete result of the tests performed in order to find the best possible configuration for the prediction accuracy of the LR algorithm is presented.

```
BEGIN

Step 1 (Loading data):

1       LOAD libraries
2       INPUT influx data to training and test model
3       SET df = change influx format data to pandas format to training model
4       SET df_2 = change influx format data to pandas format to test model
        ...

Step 2 (Preparing data to create training and test models):

        ...
5       SET tempfreq = ['15min', '30min', '45min', '1H',
6                       '1.25H', '1.5H', '1.75H', '2H']
7
8       FOR each tempfreq value
9               SET means = df median for each tempfreq value
10              SET means_2 = df_2 median for each tempfreq value
11              SET derivations = 3
12              FOR each derivations value
13                      CALL derive_nth_day_feature(means)
14                      CALL derive_nth_day_feature(means_2)
15                              SET varcorr = list(df)
16                              FOR each varcorr paramater
17                                      CALL means.corr()[varcorr]
18                                      SET predictcorr = best correlations
                                        ...
```

Figure A.1: Pseudo code of linear regression algorithm.

```
Step 3 (Creating training model):

        FOR each varcorr paramater
        ...
19              SET X = predictcorr of current varcorr paramater
20              SET y = current varcorr paramater
21
22              CAL X_train, X_test, y_train, y_test =
23               train_test_split(X, y, test_size=0.1, random_state=12)
24
25              CALL regressor.fit(X_train, y_train)
26              SET prediction = regressor.predict(X_test)
27
28              SET a1 = regressor.score(X_test, y_test)
29              SET a2 = mean_absolute_error(y_test, prediction)
                ...

Step 4 (Creating test model):

        FOR each varcorr paramater
        ...
30              SET X_test_2 = predictcorr of current varcorr paramater
31              SET y_test_2 = current varcorr paramater
32              prediction_2 = regressor.predict(X_test_2)
33              a1_2 = regressor.score(X_test_2, y_test_2)
34              a2_2 = mean_absolute_error(y_test_2, prediction_2)
                ...

Step 5 (Saving results):

        FOR each varcorr paramater
        ...
36              CALL plt.savefig()
37              CALL (a1,a2,a1_2,a2_2).to_csv()

END
```

Figure A.2: Pseudo code of linear regression algorithm.

ppmMQ135						ppmZE08					
Sampling time	D1	D2	D3	D4	Sum Time	Sampling time	D1	D2	D3	D4	Sum Time
15min	0.9967	0.9967	0.9968	0.9968	3.987	15min	0.9951	0.995	0.9936	0.9929	3.977
30min	0.9978	0.9978	0.9918	0.9981	3.985	30min	0.9976	0.9968	0.9979	0.9973	3.990
45min	0.9985	0.9987	0.9986	0.9986	3.994	45min	0.9984	0.9987	0.9985	0.9984	3.994
1H	0.9988	0.9989	0.9987	0.999	3.995	1H	0.9985	0.9989	0.9988	0.9991	3.995
1.25H	0.9992	0.9991	0.999	0.9991	3.996	1.25H	0.9992	0.9991	0.999	0.9991	3.996
1.5H	0.9994	0.9994	0.9992	0.9994	3.997	1.5H	0.9993	0.9994	0.9992	0.9994	3.997
1.75H	0.9995	0.9994	0.9992	0.9984	3.996	1.75H	0.9993	0.9994	0.9992	0.9984	3.996
2H	0.9994	0.9995	0.9995	0.9773	3.976	2H	0.9991	0.9995	0.9995	0.9807	3.979
Sum Deriv	7.989	7.990	7.983	7.967	31.928	Sum Deriv	7.986	7.987	7.986	7.965	31.924
gasBME						tempBME					
Sampling time	D1	D2	D3	D4	Sum Time	Sampling time	D1	D2	D3	D4	Sum Time
15min	0.9929	0.9982	0.9981	0.9983	3.988	15min	0.9818	0.9794	0.9796	0.9791	3.920
30min	0.9744	0.9917	0.9915	0.9899	3.946	30min	0.9454	0.9576	0.9537	0.949	3.806
45min	0.9475	0.9797	0.9778	0.9599	3.865	45min	0.9095	0.9229	0.9126	0.798	3.543
1H	0.913	0.9691	0.9576	0.8457	3.685	1H	0.8501	0.8598	0.7892	0.8173	3.316
1.25H	0.8745	0.9499	0.9074	0.8136	3.545	1.25H	0.8263	0.837	0.6665	0.8249	3.155
1.5H	0.8215	0.9196	0.6752	0.506	2.922	1.5H	0.7694	0.7222	0.532	0.6119	2.636
1.75H	0.7799	0.8823	0.6762	0.8043	3.143	1.75H	0.7597	0.6921	0,3969	0.6644	2.513
2H	0.723	0.7024	0.2986	0.4895	2.214	2H	0.6465	0.4873	0.2125	0.0902	1.437
Sum Deriv	7.027	7.393	6.482	6.407	27.309	Sum Deriv	6.689	6.458	5.443	5.735	24.325
humidBME						altBME					
Sampling time	D1	D2	D3	D4	Sum Time	Sampling time	D1	D2	D3	D4	Sum Time
15min	0.9741	0.9733	0.9728	0.9731	3.893	15min	0.9874	0.9959	0.9957	0.996	3.975
30min	0.9394	0.959	0.9558	0.9597	3.814	30min	0.9539	0.9852	0.9843	0.9838	3.907
45min	0.8901	0.9249	0.9214	0.9078	3.644	45min	0.8985	0.9705	0.9676	0.9654	3.802
1H	0.832	0.9373	0.8222	0.6545	3.246	1H	0.8382	0.9443	0.9429	0.919	3.644
1.25H	0.7562	0.7523	0.7223	0.5616	2.792	1.25H	0.7777	0.891	0.8871	0.7823	3.338
1.5H	0.6841	0.6705	0.4694	0.4764	2.300	1.5H	0.6551	0.8271	0.5576	0.618	2.658
1.75H	0.6149	0.6405	0.4255	0.5346	2.216	1.75H	0.5957	0.5278	0.3565	0.8413	2.322
2H	0.4908	0	0.4366	0.4213	1.349	2H	0.4643	0.4494	0.3911	0	1.305
Sum Deriv	6.182	5.858	5.726	5.489	23.534	Sum Deriv	6.171	6.591	6.083	6.106	24.951

Figure A.3: Complete a1 results table.