

THERMAL COMFORT OF THE ENVIRONMENT WITH INTERNET OF THINGS, BIG DATA AND MACHINE LEARNING

Matheus G. do Nascimento and Paulo B. Lopes

Graduate Program in Electrical Engineering and Computing,
Mackenzie Presbyterian University, São Paulo, Brazil

ABSTRACT

This research proposes to evaluate the level of thermal comfort of the environment in real time using Internet of Things (IoT), Big Data and Machine Learning (ML) techniques for collecting, storage, processing and analysis of the concerned information. The search for thermal comfort provides the best living and health conditions for human beings. The environment, as one of its functions, must present the climatic conditions necessary for human thermal comfort. In the research, wireless sensors are used to monitor the Heat Index, the Thermal Discomfort Index and the Temperature and Humidity Index of remote indoor environments to intelligently monitor the level of comfort and alert possible hazards to the people present. Machine learning algorithms are also used to analyse the history of stored data and formulate models capable of making predictions of the parameters of the environment to determine preventive actions or optimize the environment control for reducing energy consumption.

KEYWORDS

Big data, internet of things, machine learning, thermal comfort.

1. INTRODUCTION

People spend more than 90 % of their day in indoor environments, such as offices, schools, colleges, commercial buildings, industrial buildings and homes. The concern with analysing the internal habitats seeks to increase people's quality of life and health as, under normal conditions, the environment should not cause feelings of discomfort and stress for human beings [1], [2].

Thermal comfort is a state of well-being that the organism experiences in a given environment. It is the result of the combination of climatological variables in the environment and, in cases of discomfort, it can have a direct impact on people's health and quality of life. Each human being can make a different judgment of the comfort state of the environment due to differences in the levels of heat exchange and production [3]. These differences are related to vascular responses, subcutaneous fat thickness, individual clothing habits, eating habits, the activity that the organism was or is performing and other conditions [4]. For World Health Organization (WHO), people are sensitive to changes of up to 0.5°C/h, that is, the greater the variation, the greater the proportion of discomfort [5].

Each location has its specific climatic factors and particularities that directly influence temperature and humidity. Therefore, it is possible to evaluate the suitability of the environment to human beings with basis on these two parameters [6].

Monitoring thermal comfort is very important in a work environment as it directly influences the development of daily tasks, impacting people's productivity during the workday [4], [5], [7]. The judgment of comfort is performed through the response of each organism to the environment. Nonetheless, in huge commercial and residential buildings, typical of modern cities, the volume of data to be collected, stored and analysed in both space and time dimensions prevents the use of traditional tools.

With the use of IoT and Big Data tools, it is possible to implement a project to classify, in real time, the level of discomfort in the environment using relationships between the air temperature (°C) and the relative humidity (%) collected in the studied surroundings. Through these parameters it is possible to calculate the thermal comfort indexes to assess the environmental conditions and alert places that may generate risks to the health of exposed people [8], [9]. The use of comfort indexes helps to assess the impact of the environment on the thermal state of humans to determine personal comfort.

The devices and tools implemented in this research use the smart home concept, providing an IoT module in each chamber of the habitat. Each module has a temperature and humidity sensor on an IoT mote connected to the Wi-Fi network that will perform an HTTP POST request to a PHP script to insert the sensor readings into a LAMP server installed on the Raspberry Pi 3B+. By collecting data from IoT devices, it will be possible to monitor and generate analyses that can contribute to thermal comfort. In these circumstances, it will be necessary to use Big Data techniques to store, process and visualize the large volume of information collected in both time and space, ensuring security and privacy for the user [10], [11].

The environmental parameter collection history is used to build and train machine learning models, using the MindsDB tool, which will be directly connected to the application database, to perform the environment parameter predictions through an SQL query. MindsDB automates and simplifies the steps for model development in order to speed up the analysis of your results, reducing the need for specialist development. During the elaboration of the algorithms, it was considered that the collected data may present imperfections, inconsistencies and be of different types in order to guarantee the security, trust and privacy of the information [12], [13], [14]

This paper leverages IoT, Big Data and Automated Machine Learning (AutoML) to enable predictive environmental analysis of the internal environment location to be used in corporate environments to assess the level of comfort and productivity of its employees. It can also be used in domestic households with a focus on elderly people who need special care, being possible to identify the conditions of the environment and alert possible dangers through devices connected to the internet, physically distant from the monitored environment. By leveraging the results of predictive analysis, one can anticipate events, take actions to avoid closed environment related issues and even optimize the use of the energy in Heating, Ventilating and Air Conditioning (HVAC) systems.

2. THERMAL COMFORT

According to American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), thermal comfort is a mental condition that expresses the body's satisfaction with the environment. For the organism to remain in thermal comfort it is necessary that the heat fluxes between the environment and the organism are null [3], [15].

Metabolism is responsible for maintaining an almost constant temperature in the body in a process that, through internal biochemical reactions combining organic fuels, produces energy for the body. The human body is always exchanging heat with the environment through the skin, that

is, the skin is the thermoregulatory organ of mankind. Skin temperature depends on the blood flow that is flowing through it, and the greater the flow, the higher the temperature [4].

The American Ventilation Commission conducted a study in 1916 to analyse the influence of thermo-hygrometric conditions on labour income. The research motivation was to analyse the effects on production caused by the Industrial Revolution and extreme situations of war. With the study, it was possible to determine that an increase of the ambient temperature in the factories from 20°C to 24 °C and from 20°C to 30 °C caused worker's productivity to decrease by 15% and 28% respectively. Likewise, it was identified that with the increase in temperature in mines from 19°C to 27 °C, the miners' income decreases by 41%. Seasons impact industry production and thermally uncomfortable work environments have high incident rates [16].

The population's health and well-being are directly related to climatological variables, directly influencing people's behaviour. To analyse how much the population can suffer from thermal discomfort, the Heat Index (HI), Thermal Discomfort Index (TDI) and the Temperature and Humidity Index (THI) have been universally used [9], [17], [18] in industry, health and civil engineering. The idea of calculating comfort indexes aims to relate several variables into a single parameter.

The Heat Index is a measure to define the intensity of heat a person feels, relating the air temperature (°C) with the relative humidity (%). The Thermal Discomfort Index and the Temperature and Humidity Index are also calculated using the air temperature (°C) with the relative humidity (%). They are important parameters to determine whether the environment provides thermal comfort for people, relating them to alert levels and physiological symptoms.

To calculate the Heat Index, Equation 1 was used, in which the HI is the Heat Index in °C, the T is the temperature in °C and the UR is the relative humidity of air in % [17, p. 517].

$$\begin{aligned}
 HI = & -8.78469475556 + 1.61139411 \cdot T + 2.33854883889 \cdot UR \\
 & + (-0.14611605 \cdot T \cdot UR) + (-0.012308094 \cdot T^2) \\
 & + (-0.0164248277778 \cdot UR^2) + 0.002211732 \cdot T^2 \cdot UR \\
 & + 0.00072546 \cdot T \cdot UR^2 + (-0.000003582 \cdot T^2 \cdot UR^2)
 \end{aligned}$$

Equation 1. Heat Index Equation

In Table 1, the HI is related to danger level and physiological symptoms [17].

Table 1. Relation of HI with possible physiological symptoms

HI[°C]	Danger level	Heat syndrome
Less than 27°C	Absence of alert	Absence
From 27°C to 32°C	Warning	Possible fatigue in cases of prolonged exposure
From 32°C to 41°C	Very careful	Possibility of cramps, exhaustion and heat stroke for prolonged exposure and physical activity
From 41°C to 54°C	Danger	Cramps, heat stroke and burnout likely. Possibility of brain damage (CVA) for prolonged exposure with physical activity
Greater than 54°C	Extreme danger	Heat stroke and impending cerebrovascular accident (CVA)

To calculate the Thermal Discomfort Index, Equation 2 was used, in which the TDI is the Thermal Discomfort Index in °C, the T is the temperature in °C and the UR is the relative humidity in % [18, p. 37]. The relationship between TDI and danger level is shown in Table 2.

$$TDI = T - (0,55 - 0,0055 \cdot UR) \cdot (T - 14,5)$$

Equation 2. Thermal Discomfort Index Equation

Table 2. Thermal discomfort level as a function of the TDI

TDI[°C]	Danger level
Less than 21°C	No discomfort
From 21°C to 25°C	Less than 50% of the population feels discomfort
From 25°C to 28°C	More than 50% of the population feels uncomfortable
From 28°C to 30°C	Most of the population feels discomfort
From 30°C to 32°C	Everyone feels discomfort
Greater than 32°C	State of medical emergency

Equation 3 was used to calculate the Temperature and Humidity Index, in which the ITU is the Temperature and Humidity Index in °C, the T is the temperature in °C and the UR is the relative humidity in % [9, p. 154].

$$ITU = 0,8 \cdot T + \frac{UR \cdot T}{500}$$

Equation 3. Temperature and Humidity Index Equation

The relationship between ITU and danger levels is depicted in Table 3. For ITU values less than 21 °C the environment can be considered partially uncomfortable.

Table 3. Thermal comfort level as a function of the THI

THI[°C]	Danger level
From 21°C to 24°C	Comfortable
From 24°C to 26°C	Slightly uncomfortable
Greater than 26°C	Extremely uncomfortable

3. INTERNET OF THINGS, BIG DATA AND MACHINE LEARNING

The Internet of Things made it possible to interconnect the physical and digital world using smart devices, allowing the transmission and reception of collected information. The term "things" is used to define everyday equipment that uses hardware, software, sensors and actuators to collect parameters from the physical world so that they can be stored and processed on a local server or in the cloud. The IoT devices are limited in storage and processing due to the limited computing capacity of each device [19], [20].

The basic architecture of the IoT is composed of the physical, gateway and data processing layers. In the physical layer, all IoT devices are identified on the network and the information is collected and transmitted through the gateway, which will read and transmit the data obtained to the information management system, which is connected to the internet, making it possible to transmit the information to the cloud. The data processing layer is responsible for performing data analysis and mining, using Big Data Analytics concepts [11], [21].

Big Data is an abstract concept that emerged in 2010 and is based in the literature with the 5Vs: volume, velocity, variety, veracity and value. The first challenge is to work with the large volume of data, which is growing more and more, and it is necessary to use ways to treat this data in order to transform them into useful information [11], [16].

Data is stored for consultation and further analysis on storage servers, typically specific to IoT services. For storage, Big Data technologies are used, in addition to supporting the large volume of data, resources are offered for processing and viewing information. The servers also provide the time stamp for each stored data and visualization models using dashboards [11], [22].

The most common servers adopt relational databases that use Structured Query Language (SQL) to perform queries and manipulations on the data. The relational model presents a database composed of relations stored in tables with the respective values of each parameter. All information received is stored in a row of a table composed of specific columns for each variable. Not only SQL (NoSQL) database models are used, which offer greater performance and scalability for projects that use a large volume of information [19].

Many IoT applications are unable to handle data in real time due to computational limitations and latencies in the collection process. For real-time processing, agile software is needed, which can obtain information by means of data simultaneously. For the analysis of historical data, the information is already stored and available to use statistical techniques to obtain conclusions and relationships between the data and the context studied [22].

After performing the data analysis, it is possible to identify patterns and make some predictions of possible future results, using algorithms and statistical analysis. Cognitive analysis is present in projects that have learning mechanisms that can make decisions autonomously.

Machine learning is a data analysis method that automates the construction of analytical models, a branch of artificial intelligence based on the idea that systems can learn from data, identifying patterns with minimal human intervention. Algorithms are trained to abstract information and identify patterns in a large amount of data, to enable predictions and decisions to be made as the data is processed [23].

With a greater amount of data provided by big data, machine learning algorithms become more reliable and accessible, offering more efficient and intelligent machine learning. The learning model uses information from the dataset to formulate a model capable of performing prediction of system parameters [24], [25].

Before implementing the machine learning model, it is necessary to select and prepare the dataset to solve the proposed challenge. In a table, for example, each row represents a data point and its columns represent the characteristics that describe the data. Columns can be classified as features and can also contain labels. Features are used by the model to predict the labels, output attributes. After training, the user will provide a set of features for the model to return the expected label [12], [23], [24].

For data analysis it is necessary that the information is unique, randomized and also checked if there are trends or imbalances that could impact the training [23]. After this treatment, the data are divided into two sets, the training subset, used to train the model, and the evaluation subset, responsible for testing and refining the model. The set of algorithms to be applied varies according to the dataset attributes, the amount of data and what should be predicted [12], [24].

AutoML combines automation with machine learning, proposing to automate steps required for model development, enabling you to quickly build machine learning solutions using techniques that are easier to apply and reducing the need for specialists for development. This model proposes to streamline procedures so that it is possible to analyse their results as quickly as possible [12]. Diverse algorithms are applied to the problem and the best solution is selected, enabling the obtention of a fast result.

In recent years AutoML has become a much researched topic for addressing, in complement to traditional machine learning algorithms, applications in computer vision, data mining and natural language processing [12]. Machine learning automation only became possible due to the development of complex big data techniques on very fast computers and also due to the great demand for machine learning tools [24].

In the traditional machine learning model, the selection of the model and algorithms is done through human resources that abstract information from the available data using machine learning tools, based on professional knowledge. In addition, it is necessary to carry out resource engineering, consisting of planning and building resources from the data to abstract the maximum amount of information from the data set. Feature selection also removes unnecessary information, so models are more efficient, simpler, and interpretable. In the AutoML model these steps are automated by computer programs and are designed to be performed within a limited computational budget. Basic AutoML tools use a controller composed of an optimizer and an evaluator [12], [24], [25].

4. DEVELOPMENT

Based on the theoretical reference, it was possible to assess the level of comfort in the environment by measuring the indices of thermal comfort used. The assessment consists of judging the level of comfort in the environment, considering the level of danger and possible physiological symptoms for the individuals present.

For data collection regarding air temperature ($^{\circ}\text{C}$) and relative air humidity (%), IoT motes were used. Each one, illustrated in Figure 1, is composed of a DOIT ESP32 – ESP-WROOM-32 board connected to a Wi-Fi network, a DHT11 temperature and humidity sensor.

The temperature and humidity collection modules were installed in the rooms of a household located in the metropolitan region of São Paulo to analyse the comfort levels of the place. The house is inhabited by three adults who work remotely and share the rooms with each other. The modules were deployed in the living room, kitchen, office, bedroom and bathroom.

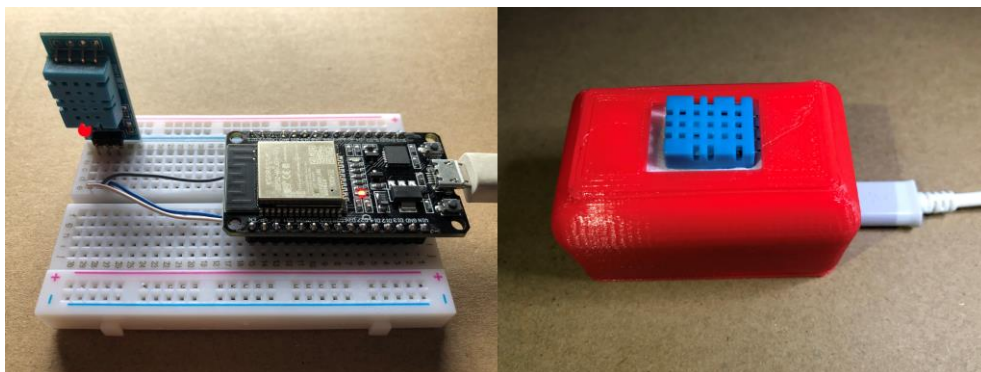


Figure 1. Prototype for collecting air temperature ($^{\circ}\text{C}$) and relative air humidity(%)

The development board DOIT ESP32 – ESP-WROOM-32 manufactured by the company DOIT (Doctors of Intelligence & Technology) uses as base the microcontroller ESP32. This component is a high-performance, low-power IoT device developed by Espressif Systems [26].

The DHT11 sensor was developed by ASAIR (Aosong Electronics Co., LTD.), includes a resistive-type HR202 humidity meter, an NTC thermistor for measuring temperature, and a high-performance 8-bit microcontroller. The temperature is measured with an accuracy of 2°C, in the range from 0°C to 50°C. Humidity is calculated with an accuracy of 5%, between 20% and 90%. DHT11 sensors do not need to be calibrated to be used because during the production process the calibration coefficients of each sensor are stored in the internal memory and are used for temperature and humidity calculations [27].

The prototype is also connected to the environment's Wi-Fi network and to the LAMP server (Linux, Apache, MariaDB and PHP) installed on the Raspberry Pi 3B+. After collecting the temperature and humidity, the DOIT ESP32 – ESP-WROOM-32 board of each environment will perform the calculation of the heat index, the thermal discomfort index and the temperature and humidity index. The apiKeyValue parameter is defined in the prototype code, which is a string used to authenticate each prototype before publishing the information in the database. The script /post-esp-data.php was implemented on the server to identify the requests sent by the prototypes and insert them into the MariaDB database of the server. Every thirty seconds an HTTP POST request is made which will send the collected information to the SensorData table in the esp_data database on the server. To display the content stored in the database, the esp-data.php script was developed, which will display all the content stored in the database through a web page when an HTTP GET request is made [28]. The SensorData table structure is specified in Figure 2.

Field	Type	Null	Key	Default	Extra
id	int(10) unsigned	NO	PRI		NULL auto_increment
sensor	varchar(30)	NO			NULL
location	varchar(30)	NO			NULL
temperature	varchar(10)	YES			NULL
humidity	varchar(10)	YES			NULL
ic	varchar(10)	YES			NULL
idt	varchar(10)	YES			NULL
itu	varchar(10)	YES			NULL
reading_time	timestamp	NO		current_timestamp()	on update current_timestamp()

Figure 2. Sensor Data structure (MariaDB)

A script /esp-data.php was developed, which will display all the content stored in the database through a web page when an HTTP GET request is made, as shown in Figure 3.

ID	Sensor	Location	Temperature	Humidity	IC	IDT	ITU	Timestamp
3549	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:23:13
3548	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:22:43
3547	DHT11	Office	23.50	64.00	24.82	23.58	21.81	2020-06-18 22:22:13
3546	DHT11	Office	23.50	64.00	24.82	23.58	21.81	2020-06-18 22:21:43
3545	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:21:13
3544	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:20:43
3543	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:20:13
3542	DHT11	Office	23.60	64.00	24.88	23.69	21.90	2020-06-18 22:19:43
3541	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:19:13
3540	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:18:42

Figure 3. esp-data.php script for viewing database content

Figure 4 presents the proposed and implemented architecture in this work. The IoT prototype collects the environment parameters and transmits them, using the Wi-Fi network, to the LAMP server installed on a Raspberry Pi 3B+ that contains a MariaDB database and two scripts. The database stores all the information collected by the IoT prototype, enabling future analysis and consultation. The script `/post-esp-data.php` is responsible for inserting the information collected by the IoT prototype and storing it in the database. The visualization can be done through the script `/esp-data.php`, also using the Grafana and MindsDB tools. The `/esp-data.php` script requests all content stored in the database, which can be used to integrate the proposed solution with other existing systems.

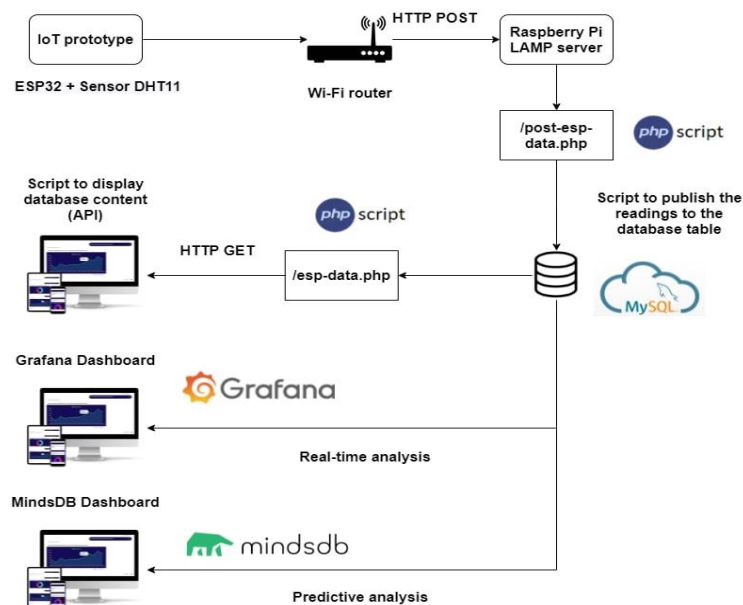


Figure 4. Project architecture

5. RESULTS

The presentation of data is carried out through the Grafana tool, which connects to the database, making it possible to consult the stored information and display them through tables, graphs and statistics. The Grafana platform is widely used for monitoring systems to generate real-time graphics using dynamic dashboards. In the tool settings it is also possible to configure alerts to notify the user in cases of abnormalities [29].

For the project, a Gmail account was created, enabled for connections from external programs to send alerts parameterized in the dashboard for each analysed environment. It was necessary to install the ssmtp package on the server, responsible for automating the transmission of e-mails via the command line so that the Grafana server triggers alerts. The dashboard was configured to average the temperature and humidity of the last fifteen minutes and trigger alerts whenever that the parameters are outside a pre-defined range. Figure 5 illustrates the Grafana dashboard and how the heat syndrome and the comfort level were determined, using the status values in each environment.

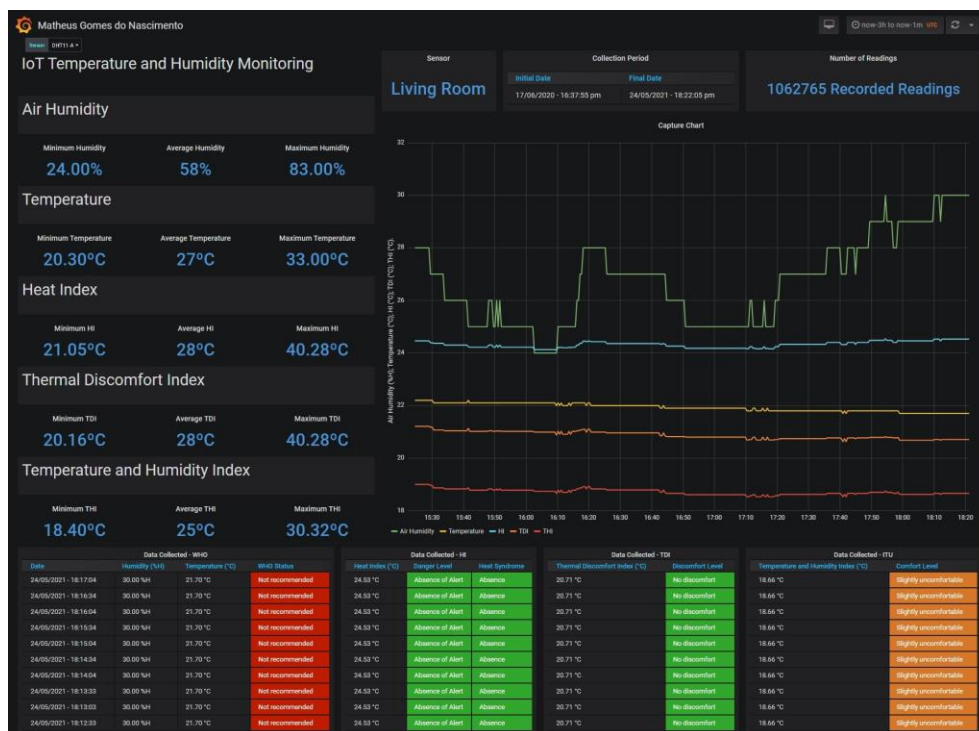


Figure 5. Grafana Dashboard containing graphs and statistics of collected data

The dashboard displays the minimum, average and maximum of each parameter as well as the graph of parameters as a function of time, according to the interval and environment defined by the user. The last data stored in each environment are also displayed through tables containing the date of collection (DD/MM/YYYY - HH:mm:ss), humidity (%), temperature (°C), the comfort indexes (°C), the heat syndrome and the status of each parameter according to Tables 4, 5 and 6.

Table 4. Relation of HI with possible physiological symptoms

HI[°C]	Danger level	Heat syndrome	Status
Less than 27°C	Absence of alert	Absence	4
From 27°C to 32°C	Warning	Possible fatigue in cases of prolonged exposure	3
From 32°C to 41°C	Very careful	Possibility of cramps, exhaustion and heat stroke for prolonged exposure and physical activity	2
From 41°C to 54°C	Danger	Cramps, heat stroke and burnout likely. Possibility of brain damage (CVA) for prolonged exposure with physical activity	1
Greater than 54°C	Extreme danger	Heat stroke and impending cerebrovascular accident (CVA)	0

Table 5. Thermal discomfort level as a function of the TDI

TDI[°C]	Danger level	Status
Less than 21°C	No discomfort	5
From 21°C to 25°C	Less than 50% of the population feels discomfort	4
From 25°C to 28°C	More than 50% of the population feels uncomfortable	3
From 28°C to 30°C	Most of the population feels discomfort	2
From 30°C to 32°C	Everyone feels discomfort	1
Greater than 32°C	State of medical emergency	0

Table 6. Thermal comfort level as a function of the THI

THI[°C]	Danger level	Status
From 21°C to 24°C	Comfortable	2
From 24°C to 26°C	Slightly uncomfortable	1
Greater than 26°C	Extremely uncomfortable	0

MindsDB is an open source tool for automating machine learning projects that allows you to quickly and efficiently build and train predictive data models using standard SQL language [13], [30].

The tool allows you to import a file with the data, connect to a BI application or connect directly to a database to create artificial intelligence tables (AI Tables) that return forecasts through an SQL query, supporting the analysis of numbers, texts, categories, timestamps and certain types of images [13].

Predictions are obtained using AutoML algorithms, ensuring accurate results in a short period of time, requiring the developer to select the data and perform an initial configuration for MindsDB to automate the entire analytical process to generate the model [13], [30].

Before training the models, a pre-processing of the data was done to prepare, organize and structure the data to optimize the final quality of the data to be analysed, consequently impacting the prediction model. The data collection was started on June 17th, 2020 and on October 20th, 2021 the SensorData table had 2038830 records with a size of 163.7 MB of storage. The first step was to remove non-numeric, null and noisy data. On April 17th, 2021, for example, the sensor installed in the office failed and collected noisy data that would compromise the quality of the data, illustrated in Figure 6.

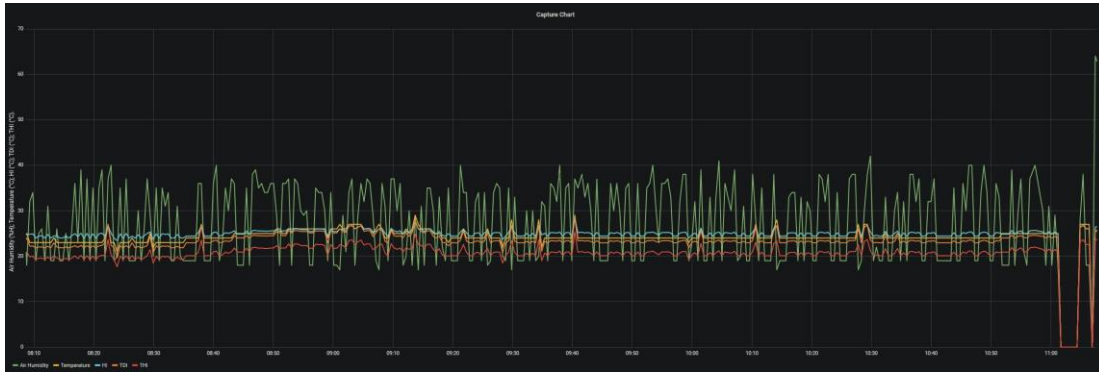


Figure 6. Noisy data collection caused by sensor failure

Data was analysed with the average of each parameter per hour to increase efficiency, reduce the amount of data and processing costs for each model. Numeric values have also been added to represent the status of each parameter, as per WHO recommendations and Tables 4, 5 and 6.

Each machine learning model implemented has its own dataset that contains only the data relevant to predicting the specific parameter in each environment. The dataset used to predict heat syndrome, for example, contains only the date, temperature, humidity, heat index and heat syndrome. Table 7 contains the machine learning models implemented with the parameters used in each dataset and the predict parameter.

Table 7. Prediction of parameters in the developed machine learning models

AutoML models	Dataset parameters	Predict parameter
who_status	reading time, humidity, temperature and sensor	WHO status
hi_danger_level_heat_syndrome	reading time, humidity, temperature, IC and sensor	Danger level and heat syndrome
tdi_discomfort_level	reading time, humidity, temperature, IDT and sensor	Discomfort level
thi_comfort_level	reading time, humidity, temperature, ITU and sensor	Comfort level
humidity_temperature	reading time, humidity, temperature and sensor	Humidity and temperature

The who_status model uses Logistic Regression, k-Nearest Neighbors, Decision Trees, Support Vector Machine and Naive Bayes algorithms. For the other models that use multiclass classification, the tool uses the K-Nearest Neighbors, Decision Trees, Naive Bayes, Random forest and Gradient boosting algorithms. The models are trained using the different algorithms to assess which is the most efficient, using the time required for training and the accuracy to make the correct predictions [13]. Table 8 specifies the models implemented in each environment.

Table 8. Machine learning models developed for each environment.

Location	Sensor	AutoML template
Living room	DHT11-A	who_status_a, hi_danger_level_heat_syndrome_a, tdi_discomfort_level_a, thi_comfort_level_a, humidity_temperature_a
Kitchen	DHT11-B	who_status_b, hi_danger_level_heat_syndrome_b, tdi_discomfort_level_b, thi_comfort_level_b,

		humidity_temperature_b
Bedroom	DHT11-C	who_status_c, hi_danger_level_heat_syndrome_c, tdi_discomfort_level_c, thi_comfort_level_c, humidity_temperature_c
Bathroom	DHT11-D	who_status_d, hi_danger_level_heat_syndrome_d, tdi_discomfort_level_d, thi_comfort_level_d, humidity_temperature_d
Office	DHT11-F	who_status_f, hi_danger_level_heat_syndrome_f, tdi_discomfort_level_f, thi_comfort_level_f, humidity_temperature_f

To train the models in MindsDB, advanced configuration parameters using Raspberry Pi 3 B+ Graphics Processing Unit (GPU) were used. For advanced parameter configuration, the tool is informed that the dataset lines are sequentially associated by the reading time feature, the target parameters for each model, the number of previous samples that will be used to perform the forecast and the error margin of the sample set. In this project, 24 previous samples were verified and the margin of error of the sample set of 2° C for temperature and 5% for humidity, considering the use of the DHT11 sensor.

After calculating the predictor model, a report is generated specifying the number of samples used for the training and testing phase of the model, the relevance of each feature to the model, the confusion matrix and the model's accuracy. Figure 7 present the report generated in MindsDB after calculating the predictor model of humidity in the living room.

The confusion matrix is an NxN matrix that allows extracting metrics to assist in the evaluation of classification models, with N being the number of target classes. The matrix compares the actual data with the data predicted by the model and it is possible to identify the model's performance and the probability of success and error for each target class [13].

Accuracy describes how often the machine learning model correctly classifies a data point, that is, the rate at which the model makes a correct prediction. The greater the accuracy of an algorithm, the lower the risk for making decisions using the model. Precision is obtained by summing the number of times the model predicted the true positives and true negatives by the total number of predictions. Table 9 shows the accuracy of each predictive model calculated in each analysed environment [24].

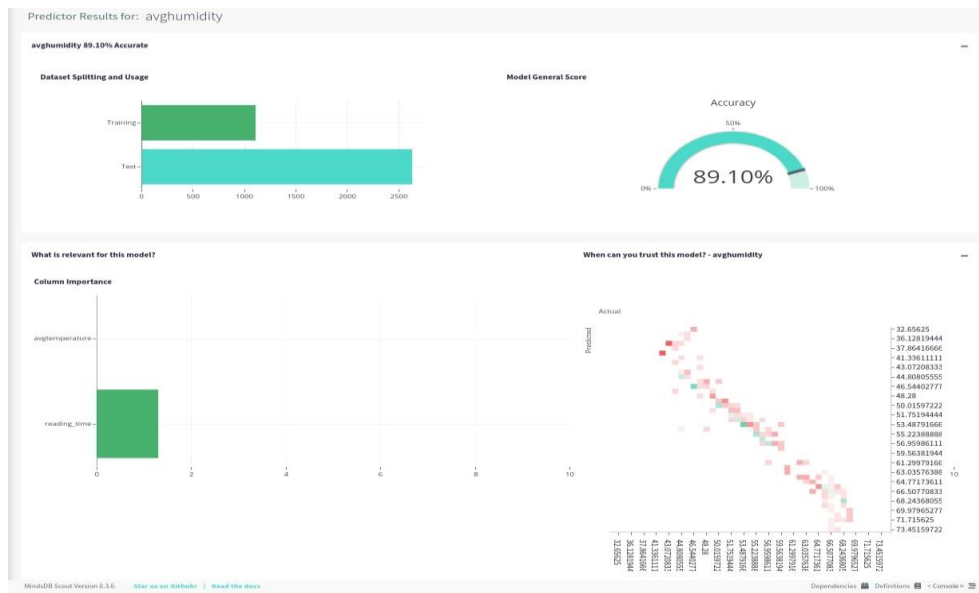


Figure 7. Results of the predictor model of the average humidity in the living room

Table 9. Machine learning models developed for each environment

Location	Target parameter	Accuracy
Living room	Who Status	93,8%
	Danger level and heat syndrome	92,1%
	Discomfort level	77,3%
	Comfort level	100%
	Relative air humidity (%)	89,1%
	Air temperature (°C)	64,5%
Kitchen	Who Status	99,6%
	Danger level and heat syndrome	86,9%
	Discomfort level	99,5%
	Comfort level	96,7%
	Relative air humidity (%)	76,6%
	Air temperature (°C)	76,0%
Bedroom	Who Status	91,6%
	Danger level and heat syndrome	100%
	Discomfort level	100%
	Comfort level	77,8%
	Relative air humidity (%)	90,4%
	Air temperature (°C)	80,9%
Bathroom	Who Status	98,3%
	Danger level and heat syndrome	100%
	Discomfort level	100%
	Comfort level	90,9%
	Relative air humidity (%)	48,0%
	Air temperature (°C)	81,9%
Office	Who Status	100%
	Danger level and heat syndrome	100%
	Discomfort level	65,3%
	Comfort level	100%
	Relative air humidity (%)	72,9%
	Air temperature (°C)	60,5%

The `who_status_a` model has positive results for uncomfortable environments according to the WHO recommendations, with numeric status 0, and negative results for comfortable places, with numeric status 1. During training, the relationships and the importance of each feature in the dataset, with moisture being classified as the most relevant parameter for the model, with an importance of 4.40 on a scale of 10.00. The confusion matrix of this model correctly predicted 99.00% of positive results and 93.00% of negative results with an overall accuracy of 93.80%.

During the capture period, the `hi_danger_level_heat_syndrome_a` model registered only the environment with no danger level and heat syndrome, with numerical status 4, and the environment in a state of attention with possible fatigue in cases of prolonged exposure, with numerical status 3. Temperature was rated as the most relevant parameter for the model, with importance 1.50 on a scale of 10.00. This model's confusion matrix correctly predicted 100.00% of numeric 3 statuses and 98.00% of numeric 4 statuses with an overall precision of 92.10%.

The `tdi_discomfort_level_a` registered only the environment without discomfort, with the numeric status 5, and the uncomfortable environment for less than half of the people, with the numeric status 4. The confusion matrix of this model correctly predicted 95.00% of the numeric status 4 and 98.00% of numeric status 5 with an overall precision of 77.30%.

The `thi_comfort_level_a` model registered the comfortable environment, with status 2, the slightly uncomfortable environment, with status 3, and the extremely uncomfortable environment, with status 0. Temperature was classified as the most relevant parameter for the model, with importance 5.00 on a scale of 10.00. The confusion matrix of this model correctly predicted 100.00% of the samples in different statuses with an overall accuracy of 100.00%.

The `humidity_temperature_a` model recorded the average air humidity between 44.00% and 83.00% and the average ambient temperature between 19.30°C and 33.00 °C. The time of collection was rated as the most relevant parameter for the model with importance 1.30 and 6.40 on a scale of 10.00 for humidity and temperature respectively. The overall accuracy for forecasting humidity is 89.10% and for temperature 64.50%.

Detailed predictive analytics of environmental parameters in the kitchen, bedroom, bathroom and office can be achieved using the reports generated in MindsDB and the same procedure described for the living room.

Figure 8 contains a comparison between actual and predicted values of temperature and humidity in the living room as of September 6th, 2021. The average difference between actual and predicted moisture was 1.77%, with a maximum variation of 6.09%. For temperature, the average difference was 0.67°C with the maximum variation of 2.54°C. In this context, it is important to note that this predictive analysis was performed in an environment with no barriers to protect from external temperature or humidity sudden variations. This is the main reason why there is some deviation from observed actual data.

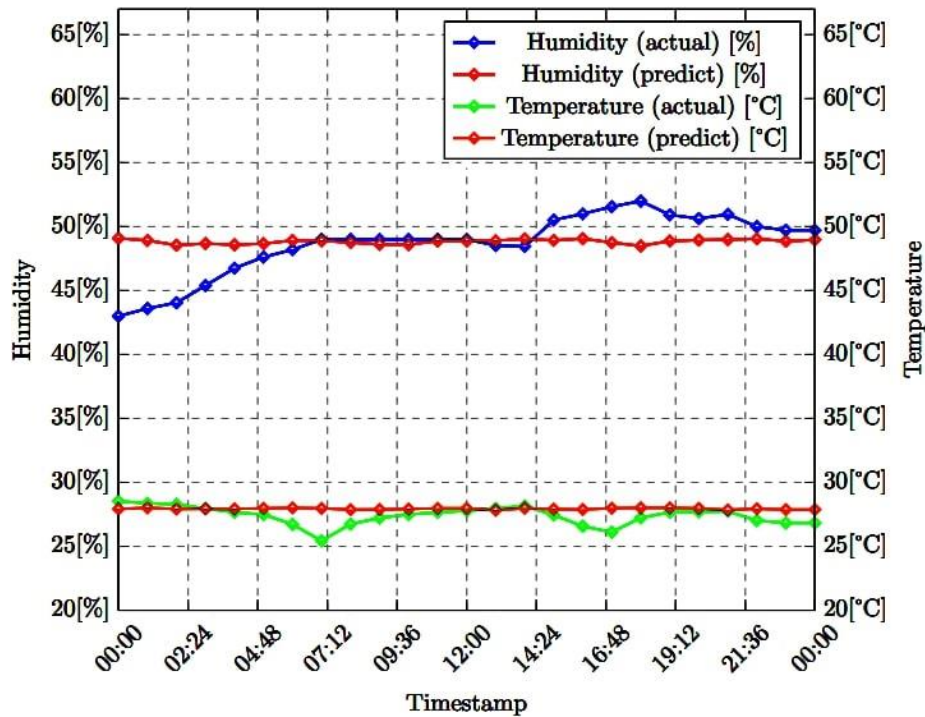


Figure 8. Comparison between predicted and actual humidity/temperature on September 6th, 2021

6. CONCLUSION

Data analysis becomes more complex as the amount of stored information increases, making it necessary to use elements of visual communication to assess the comfort of the environment in an intuitive and simple way.

The tools used promote the traceability and organized visualization of the information collected in each environment, synthesizing the relevant parameters in graphics, indicators and alerts so that the user can understand the results obtained.

Comfort in an indoor environment is analysed using relationships between air temperature (°C) and relative humidity (%) to assess environmental conditions and alert places that may pose risks to the health of exposed people, using the indices of thermal comfort.

The results of the environmental analysis are illustrated using the Grafana dashboard to assess comfort in real time and the MindsDB dashboard to predict the variables related to thermal comfort, including the comfort level and possible physiological causes for an individual who remains in the environment.

Machine Learning prediction was performed and the obtained data indicate that it is possible to reliably anticipate the evolution of the thermal comfort in internal environments. Therefore, it is possible to both take preventive actions to avoid future deviations of temperature and humidity as well adopt energy profiles to reduce the power consumption in a ecological-friendly system.

For future work, the number of IoT devices will be increased and the current LAMP server replaced by a cloud server to ensure there is no reduction in performance while increasing

demand for the server. Thus, server resources will be increased during the calculation of machine learning models and also scaled according to the number of devices and sensors.

REFERENCES

- [1] A. Datta, R. Suresh, A. Gupta, D. Singh, P. Kulshrestha, "Indoor air quality of non-residential urban buildings in Delhi, India" *Int. Journal of Sustainable Built Environment*, vol. 6, no. 2, pp. 412, Dec. 2017, DOI: 10.1016/j.ijsbe.2017.07.005.
- [2] F. H. Mohamed Salleh, M. B. Saripuddin and R. bin Omar, "Predicting thermal comfort of HVAC building using 6 thermal factors," in *2020 8th Int. Conf. on Information Technology and Multimedia (ICIMU)*, Selangor, Malaysia, 2020, pp. 170–175. DOI: 10.1109/ICIMU49871.2020.9243466.
- [3] J. K. Page, "The thermal environment," "The air quality environment," in *World Health Organization - Environmental Health in Rural and Urban Development and Housing Unit*, 1st ed. Geneva, Geneva, Switzerland: WHO, 1990, ch. 1 and 2, pp. 15–64. [Online]. Available: <https://apps.who.int/iris/handle/10665/62723>.
- [4] K. Fabbri, "Ergonomics of the thermal environment. Human body and clothes," in *Indoor Thermal Comfort Perception*, 1st ed, Cesena, Italy: Springer, Cham, 2015, ch. 3, sec. 1, pp. 25–63. DOI: 10.1007/978-3-319-18651-1.
- [5] G. Gao, J. Li and Y. Wen, "DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8472–8484, Sep. 2020, DOI: 10.1109/JIOT.2020.2992117.
- [6] K. Nkurikiyeyezu, "An efficient thermal comfort delivery in work- places," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kyoto, Japan, 2019, pp. 427–428. DOI: 10.1109/PERCOMW.2019.8730680.
- [7] M. R. Alam, M. B. I. Reaz, M. A. M. Ali, "A review of smart homes: past, present, and future," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1190–1203, Nov., 2012, DOI: 10.1109/TSMCC.2012.2189204.
- [8] H. Singh, V. Pallagani, V. Khandelwal and U. Venkanna, "IoT based smart home automation system using sensor node," in *2018 4th Int. Conf. on Recent Advances in Information Technology (RAIT)*, Dhanbad, India, 2018, pp. 1–5. DOI: 10.1109/RAIT.2018.8389037.
- [9] G.M. Barbirato, L. C. L. Souza and S. C. Torres, "Clima e planejamento urbano: por que é importante obter informações sobre o clima," in *Clima e cidade: a abordagem climática como subsídio para estudos urbanos*, 1st ed. Maceió, AL, BR: EDUFAL, 2007, ch. 6, pp. 133–148.
- [10] E. N. Barboza, C. R. A. Caiana, F. C. B Neto, A. G. Maia, C. J. L. Lima, P. B. Maracaja, "Análise do índice de calor (IC), índice de conforto térmico (IDT) e índice de temperatura e umidade (ITU) na cidade de Iguatu/CE a partir de dados históricos," *Revista Brasileira de Gestão Ambiental*, vol. 13, no. 3, pp. 2–5, Jul, 2019.
- [11] M. A. Borges, P. B. Lopes, L. A. Silva, M. de O. Igarashi, G. M. F. Correia. "An architecture for the internet of things and the use of big data techniques in the analysis of carbon monoxide," in *2017 IEEE International Conf. on Inf. Reuse and Integration (IRI)*, San Diego, USA, 2017, pp. 184–191.
- [12] B. P. Santos, L. A. Silva, C. S. Celes, J. B. Borges, B. S. P. Neto, M. A. M. Vieira, L. F. M. Vieira, O. N. Goussevskaia, A. A. Loureiro, "Internet das coisas: da teoria a prática," in *Minicursos SBRC - Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Porto Alegre, Brazil, 2016, pp. 3–27.
- [13] I. Guyon, Y. Hu, Y. Li, W. Tu, Q. Yang, Y. Yu, "Taking human out of learning applications: A survey on automated machine learning," *ArXiv*, vol. abs/1810.13306, pp. 1–18, Dec, 2019. Accessed on: June, 21, 2021, [Online]. Available: <https://arxiv.org/pdf/1810.13306.pdf>.
- [14] P. Backman, "Deploy machine learning models directly in your database with AI Tables," in *MindsDB Documentation*. Los Angeles, CA, USA: MindsDB, 2020, pp. 1–6. [Online]. Available: <https://mindsdb.com/wp-content/uploads/2020/12/MindsDB-AI-Tables-Whitepaper-2020.pdf>.
- [15] American Society of Heating, "Thermal comfort," in *ASHRAE Handbook Fundamentals*, Inch-Pound ed., Atlanta, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2009, ch. 9, sec. 1.
- [16] Y. Demchenko, C. Ngo, C. Laat, P. Membrey, D. Gordijenko, "Big security for big data: Addressing security challenges for the big data infrastructure," in *10th VLDB Workshop on Secure Data Management*, vol. 2, Cham, Switzerland, 2014, pp. 3–5. DOI:10.1007/978-3-319-06811-413.

- [17] K. Blazejczyk, Y. Epstein, G. Jendritzky, H. Staiger, B. Tinz, "Comparison of UTCI to selected thermal indices" *Int. journal of biometeorology*, vol. 56, no. 3, pp. 515-517, May 2011, DOI: 10.1007/s00484-011-0453-2.
- [18] T. A. Yousif, H. M. M. Tahir, "Application of Thom's thermal discomfort index in Khartoum state, Sudan" *J Forest ProdIndust*, vol. 2, no. 5, pp. 36-38, Sep. 2013.
- [19] S. Greenland, "The internet of changes everything," in *The Internet of Things, 1st ed.* Cambridge, Massachusetts, USA: MIT Press, 2015, ch. 1, 2, 4 and 5, pp. 11-24 and 33-47.
- [20] H. Garg, M. Dave, "Securing IoT devices and securely connecting the Dots using REST API and middleware," presented at the *4th Int. Conf. on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, Ghaziabad, India, Jul. 29, 2019. DOI: 10.1109/IoT-SIU.2019.8777627.
- [21] C. Savaglio, G. Fortino, M. Zhou, "Towards interoperable, cognitive and autonomic IoT systems: An agent-based approach," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, Reston, VA, USA, 2016, pp. 58-59. DOI:10.1109/WF-IoT.2016.7845459.
- [22] R. R. F. Dias, "IoT sem mistérios," in *Internet das Coisas sem mistérios: Uma nova inteligência para os negócios, 1st ed.* City of Publisher, São Paulo, Brazil: Netpress Books, 2016, ch. 4 and 5, pp. 51-54 and 71-92.
- [23] IBM Cloud Education, "What is machine learning?," in IBM Cloud Education, Armonk, New York, USA: IBM Cloud Learn Hub, 2020. [Online]. Available: <https://www.ibm.com/cloud/learn/machine-learning>.
- [24] L. G. Serrano, "What is machine learning? It is common sense except done by a computer," in *Grokking Machine Learning, 1st ed.* Greenwich, Connecticut, USA: Manning Publications, 2021, ch. 1. [Online]. Available: <https://livebook.manning.com/book/grokking-machine-learning>.
- [25] J. Waring, C. Lindvall, R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 1, no. 10, pp. 1-5, Feb., 2020, DOI. 10.1016/j.artmed.2020.101822.
- [26] *ESP32-WROOM-32 Datasheet, 3.7 ed.*, Espressif Systems, Pudong, SH, China: 2021, pp. 8-48.
- [27] *DHT11 - Humidity and Temperature Sensor, 1st ed.*, SUNROM Technologies, Isanpur, AMD, India: 2012, pp. 1-7.
- [28] R. Santos, "ESP32/ESP8266 Insert Data into MySQL Database using PHP and Arduino IDE"," *Random Nerd Tutorials*. Accessed: Oct. 25, 2020. [Online]. Available: <https://randomnerdtutorials.com/esp32-esp8266-mysql-database-php>.
- [29] Grafana LABS, "What is Grafana?," in *Grafana Documentation*. New York, NY, USA: Grafana LABS, 2020. [Online]. Available: <https://grafana.com/docs/grafana/latest/getting-started/what-is-grafana/>.
- [30] G. Leopold, "MindsDB, AutoML Startup, Gains Seed Fund- ing"," *Datanami*. Accessed: Oct. 8, 2020. [Online]. Available: <https://www.datanami.com/2020/04/20/mindsdb-automl-startup-gains-seed-funding/>.

AUTHORS

Matheus G. do Nascimento was born in São Paulo, Brazil, in 1997. He received the B. Sc. degree in electrical engineering from the Mackenzie Presbyterian University, São Paulo, Brazil, in 2019.

He is currently a scholarship student in the Graduate Program in Electrical Engineering and Computing (PPGEEC), Mackenzie Presbyterian University, São Paulo, Brazil. His current research interests include big data, internet of things, machine learning and thermal comfort.

From 2018 to 2019 he worked at Valid S.A., a Brazilian company, working as an RFID intern in the IoT area. In 2020 he was hired as a new product analyst, developing and implementing IoT and RFID solutions.



Paulo B. Lopes has received the B.Sc. and M.Sc. in EE from Federal University of Rio de Janeiro, Brazil, in 1978 and 1981, respectively, and the Ph.D. in EE from Concordia University, Montreal, Canada, in 1985.

From 1985 to 1988, he was with Elebra and CMA, two Brazilian companies, working on the design of several communication equipments. From 1988 to 1999, he was with Texas Instruments as a DSP specialist. In 1999, he moved to Motorola-SPS (later to become Freescale Semiconductor) as a Sales and Application Manager.

Since 2009, he has been with Mackenzie Presbyterian University as a professor in the School of Engineering. His research interests are Circuit Theory, Digital Signal Processing, Analog Circuit Design and Communication Theory and Internet of Things.

