Comput Syst Sci & Eng (2019) 5: 297–303

International Journal of
Computer Systems
Science & Engineering

Implementation of Decision Trees as an Alternative for the Support in the Decision-Making Within an Intelligent System in Order to Automatize the Regulation of the VOCs in Non-Industrial Inside Environments

Liliana Aguilar, Salvador W. Nava-Díaz* and Gabriel Chavira*†**

* *Facultad de Ingeniería "Arturo Narro Siller", Universidad Autónoma de Tamaulipas, Tampico, Tamps. México*

Natural ventilation is a component that provides a positive impact in the quality of air conditions in indoor environments, especially in non-industrial buildings. The maintenance of a continuous entrance of outside air through windows provides to the indoor a feasible and affordable manner to regulate and sustain low standards in the VOC (Volatile Organic Compounds). The technology and the Human Computer-Interaction have contributed to the creation of Intelligent Environments (EI) that provides to humans being a positive and non-intrusive responsiveness of the environment to improve their quality of life in daily activities. The Decision Trees for Decision Making is a mathematical analysis suitable for taking decision. This method will provide an intelligent system with the existent variables in the context that establish the requirement of the natural ventilation. The present paper shows the use of the Decision Trees as an analytical method for the decision making that can be apply in an intelligent system, in the automatize of the natural ventilation in a non-industrial closed environment. This method allows the incorporation of outside air and regulate in a significant manner the Volatile Organic Compounds presents in any occupied building. It was found that the application of Decision Trees and Shannon Entropies provide a feasible procedure for the diagnose of a real backdrop that enable the creation of a routing path for the decision making through the application the computation technology.

Keywords: VOCs, Intelligent Environments (IE), Tree Decisions, Shannon Entropies, Sick Syndrome Building (SSB)

1. INTRODUCTION

Air is an essential component for human beings. Currently, environmental pollution generates a problem of health and comfort in human beings since it is present inside buildings too. Studies of the United States Environmental Protection Agency (EPA) about the exposures that humans have with inside air pollutants indicate that many of them are 2 to 10 times more that outside air pollutants [1]. This situation aggravates because according to the World Health Organization [2] the human being remains indoors more than 80% of his

*Email: laguilar@docentes.uat.edu.mx, snava@docentes.uat.edu.mx,
†gchavira@docentes.uat.edu.mx

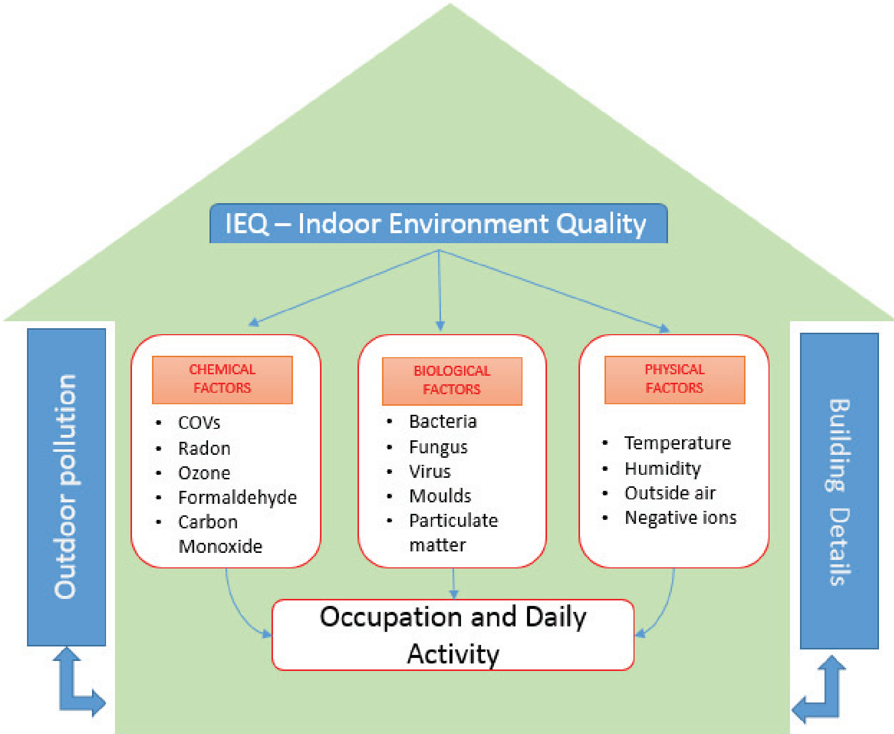

Figure 1 General Diagram IEQ (Own elaboration)

time. It also establishes statistics greater than 3.8 million of premature yearly deaths attributable to the exposure of polluted air [2].

Inside the closed spaces exist and coexist pollutants generated by indoor pollution. Indoor Pollutants produced by outside pollution, daily human beings activities, construction materials, building design, cleaning and remodeling materials, furniture and equipment, etc (Figure 1). These pollutants cause a bad IEQ (Indoor Environment Quality). One of these pollutants are the Volatile Organic Compounds (VOCs). The VOCs are a group of compounds belonging to different chemical families (alcohols, aldehydes, ketones, glycol ethers, terpenes, etc.) that have in common their carbon chemical basis and the feature of evaporate at ordinary room temperature. They are issued by various sources whether pyogenic (naturally) or anthropogenic (human-made) at outdoor and indoor. In indoor environments they are present in various items like decorative and construction materials (varnishes, paints, cleaners, glues, wood, carpets, fabrics, air fresheners, biocides) [3] and through the burning of fuels like gas, wood, coal and natural gas [4]. The VOCs are one of the elements present in the symptomatology of the Sick Syndrome Building (SSB) [5]. The SSB is characterized by a series of physical discomforts that occupants present and they disappear once they leave the building. The WHO differences two kinds of sick buildings; the temporarily sick that include the new or recent remodeling buildings where the symptoms usually disappear in half a year and the permanently sick where the symptoms persists for years. The last one are characterized for being hermetic in which the windows are not usually present and they have a forced ventilation system of common air. The main source of air inside the building is the artificial air conditioning that maintains a homogenous thermal environment (heath or cold).

The regulation of the indoor VOCs can be established through an adequate natural ventilation that allows the entrance of outdoor air when the occupant is not in the building and the environmental conditions are optimal. Computer technology is a viable tool to achieve this purpose by being able to establish Intelligent Environments (EI) with automation systems that identify the favorable context when to execute the natural ventilation when the occupant is not in the building. The context awareness will provide the system the necessary information to take the decision for the opening of the windows with the decision tree.

1.1 Related Works

Some researchers have proposed a solution to monitor the IEQ using wireless sensor networks. Such as the authors of the article [6] design a system that monitors the indoor and outside air quality inside the Microsoft Central Offices in China. It can be seen in Figure 2. The system registers the air quality in all the areas inside the building and send the information to the cloud. The cloud storages the information in a cloud-database and can be accessed through a mobile application or the website, the user may decide to stay or leave the area, or to apply an action in order to decrease the bad quality air.

Hui Xie et al. exploited the Artificial Neural Network (ANN) in order to predict the Indoor Air Quality. The results were compared to other linear regression models and it was concluded that ANN gave better predictions. Similar to Hui Xie results, in the paper [7], the authors developed an ANN air quality monitoring system with the additional function of classifying sources influencing the IEQ. Basically, the system consists of sensor module cloud (SMC), base station and service-oriented

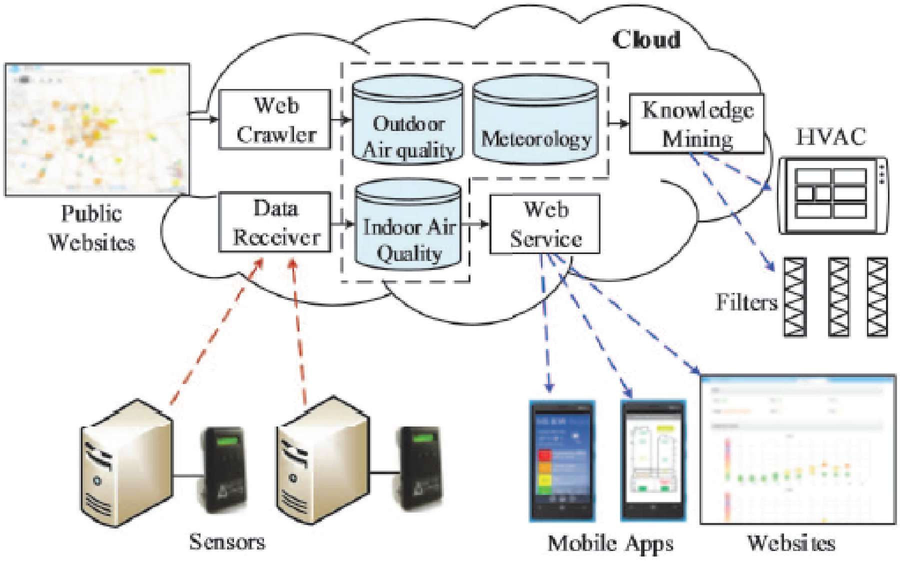

Figure 2 Architecture of the system.

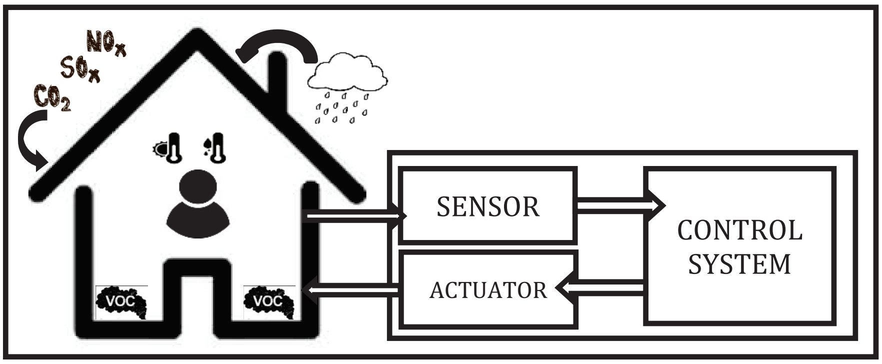

Figure 3 Technological Proposal for the automatize of the Natural Ventilation (Own elaboration)

client. The SMC contain collections of sensor modules that measure the air quality data and transmit the captured data to base station through wireless network. The IAQ monitoring system is also equipped with IAQ Index and thermal comfort index, which could tell the users about the room's conditions. The results showed that the system is able to measure the level of air quality and successfully classify the sources influencing IAQ in various environments like ambient air, chemical presence, fragrance presence, foods and beverages and human activity.

Even with the above, it can be observed that decision-making techniques directly related to the regulation of air quality in closed non-industrial spaces have not been applied, when this can have a positive impact on the IEQ through the automatize of natural ventilation.

2. PROPOSAL

The technological advances of the last decades have provided the human beings a new way of interaction with the real world and all the different devices with which they interact. The Human-Computer Interaction (HCI) has evolved in such a way that allows the implementation of Intelligent Environments (IE). The IE are physical spaces where information and ubiquitous technologies are combined for the user (human being) taking the environment of the real world as a reference. This type of environments interact with the physical worlds, from which the information is captured and processed for the taking decision that generate actions in the real world [8]. As a proposal for the automation of the natural ventilation is the implementation of an Intelligent Environment that allows the regulation of the VOCs in a building when the occupants are not inside (Figure 3). The system could be composed by a Wireless Network Sensor (WNS) for monitoring and recording information of the environment through sensors and actuators.The sensors that the system needs are of external pollution (MQ-135 from Arduino), presence of humans (Sensor HC-SR501), presence of rain (YL-83), temperature and humidity (sensor DHT22) and presence of VOCs (Sensor CCS811 – VOC/Eco2). With this information the system decides to ventilate the building by the automated opening of the windows. The present paper exposes the Decision Tree Algorithm as a methodology for the evaluation and taking decision of the system.

2.1 Methodology

Due to the importance that natural ventilation has in the Indoor Air Quality and in the regulation of some dangerous compounds present in non-industrial buildings, the present paper proposes

Table 1 Variable values.

VARIABLE	VALUES	INTERPRETATION
OP	YES / NO	YES: Outdoor pollution higher than 100 ppm NO: Good air quality (less than 100 PPM)
R	YES / NO	YES: Presence of rain. NO: Free of rain.
T	HIGH/NORMAL/LOW	HIGH: Temperature higher than 30^{o}C NORMAL: Temperature between 18^{o}C and 30^{o}C LOW: Temperature less than 18^{o}C
H	HIGH/NORMAL/LOW	HIGH: Humidity higher than 50% NORMAL: Humidity between 30% and 50% LOW: Humidity less than 30%
VOCs	HIGH/NORMAL	HIGH: Higher than 25mg/m3 NORMAL: Lessthan 25 mg/m3

Table 2 Scenario.

NUMBER	OP	R	T	H	VOC	DECISION
1	NO	NO	HIGH	HIGH	HIGH	YES
2	NO	YES	HIGH	HIGH	HIGH	NO
3	YES	NO	NORMAL	HIGH	NORMAL	NO
4	YES	NO	LOW	NORMAL	HIGH	NO
5	NO	NO	NORMAL	NORMAL	HIGH	YES
6	NO	NO	LOW	NORMAL	NORMAL	YES
7	NO	YES	NORMAL	NORMAL	NORMAL	NO
8	YES	YES	LOW	NORMAL	NORMAL	NO
9	NO	NO	HIGH	NORMAL	HIGH	YES
10	NO	NO	HIGH	LOW	NORMAL	YES
11	NO	NO	NORMAL	HIGH	HIGH	YES
12	YES	YES	HIGH	HIGH	HIGH	NO
13	YES	NO	HIGH	NORMAL	NORMAL	NO

the implementation of the Decision Trees as an alternative for the support in the decision-making within an Intelligent System in order to automatize the regulation of the VOCs in non-industrial inside environments, through the automatize of the natural ventilation. The scenario to analyze is a building that will has the next variables and their values (Table 1):

- Outdoor pollution (OP): The OP affects in the decision of the opening of the windows. An OP higher than 100 ppm is considering bad according to the National Ambient Air Quality Standards of the EPA [1]. If the OP is higher of 100 mmp the best decision is that the building remains close to avoid the entrance of harmful elements.

- Rain (R): The rain is obviously an element to consider in the opening of the windows because water can cause damages inside.

- Temperature (T): The VOCs volatilizes at ordinary room temperature, polluting the building at this process.

- Humidity (H): Humidity is also an important element for the proliferation of the VOCs. The normal standards are between the 30 and 50 per cent of relative humidity. A higher or lower humidity not only causes discomfort on human beings, also increase the VOCs in the environment [9].

- Volatile Organic Compounds (VOCs): The continuous presence of more than 25 mg/m3 of VOCs inside generate a health risk and negative symptomatology on human beings [10].

Each variable contains the values established in the next Table 1:

3. SCENARIO

With the information and interpretation of the variables exposed in the Table 1 a scenario with 13 possible real environmental conditions is built that include the variables and a Decision Column. This final column displays a YES or NO because of the analysis of the inferred values in a real context of all the variables involved (Table 2). The values of the variables are entered for

Table 3 Entropies for each variable.

OP	
YES	0
NO	0.81127812
INF-NODE	0.49924808
ENTROPY	0.49647938

R	
YES	0
NO	0.91829583
INF-NODE	0.63574327
ENTROPY	0.35998418

H	
HIGH	0.97095059
REGULA	0.98522814
LOW	0
INF-NODE	0.90394999
ENTROPY	0.09177746

T	
HIGH	1
REGULAR	1
LOW	0.91829583
INF-NODE	0.98114519
ENTROPY	0.01458226

COV	
HIGH	0.98522814
REGULAR	0.91829583
INF-NODE	0.9543363
ENTROPY	0.04139115

ENTROPIES	
OP	0.4964794
R	0.3599842
H	0.0917775
T	0.0145823
COV	0.0413911

Table 4 Entropies ordered from highest to lowest.

ENTROPIES	
OP	0.49647938
LL	0.35998418
H	0.09177746
COV	0.04139115
T	0.01458226

the composition of a table that shows with real data the decision to the opening or closing of the windows in a non-industrial building. This decision will provide fresh air of outside through the natural ventilation when the occupant is not inside.

This scenario contains a series of real events in which for it diagnose is used the Shannon Entropies. The formula quantifies the entropy of each variable:

$$f\left(P\left(X_k\right)\right) = -P\left(X_k\right)\log_2\left(\left(P\left(X_k\right)\right), \quad 0 \leq P\left(X_k\right) \leq 1\right.$$

Where $P\left(X_k\right)$ is the probability that a random variable X takes the value of k, with $k = 1, 2, \ldots, N$. Therefore, the function $f\left(P\left(X_k\right)\right)$ is equal cero when $P\left(X_k\right) = 0$ and $P\left(X_k\right) = 1$, and positive for in-between numbers, being its maximum value $\left(f\left(P\left(X_k\right)\right)\right) = 1$ when $P\left(X_k\right) = 0.5$. Who from this expression, the Shannon entropy is defined by:

$$H\left(X_k\right) = \sum_k f\left(P\left(X_k\right)\right)$$

Consequently, the entropy of a set of data is interpreted as the degree of information that provide its observation. The greater its uncertainty the greater is its entropy. The implementation of Shannon's Entropies in the variables are shown on table 3:

Each table shows each variable with the entropies values of each attribute, the general entropy and the gain. The entropies values are established using the Shannon Entropy formula, while the Gain is the subtraction of the tree gain whose value is 0.995772 from the entropy of each variable. The quantified and ordered entropies (Table 4) allow the construction of a Decision Tree.

The values of the entropies of each ordered variable and the value of the entropies of each attribute provides the information for the construction of the routing path represented by the figure 4. It can be observed that the variables that determine in an important manner the opening of the windows are at first the Outdoor Pollution and the Rain.The presence of both elements eliminate the decision to open the windows. The summation of both entropies conforms almost an 86% (.8564) and concluding that their presence are very important to maintain the windows closed.

Subsequently humidity shows a higher possibility of the presence of VOCs in its normal and high attributes. In other hand, the variable VOCs presents a high entropy on both of their attributes (Normal and High). At the end of the routing path, the temperature in its high and normal attributes represent the feasibility that natural ventilation facilitates the regulation of the VOCs.

4. VALIDATION

With the random function, a 100 hundred scenario was built. The function was applied to each variable in order to create the scenario. With the obtained values, the routing path of the decision tree is tested. In the same way, this scenario is presented to a PhD in Environment Science, who choose when the ventilation is necessary and feasible. Both decision are shown in the last 2 columns of the Table 4.

The table 5 shows how the expert decides the natural ventilation in 24 times, while the routing path decides only in 17 times. The 17 times concur with the expert. This means a 70.83% of accuracy. They are two aspects that the expert visualizes in order to open windows, the presence of outside

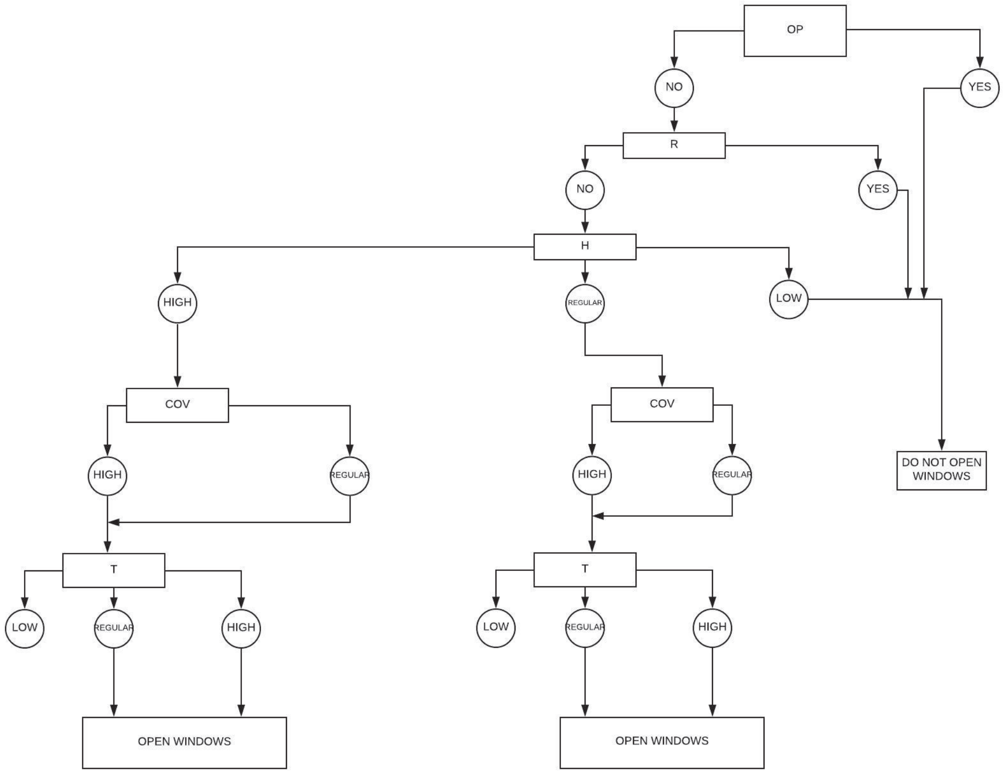

Figure 4 Tree Decision

Table 5 Decision Tree versus Expert.

OP	R	H	covs	T	DECISION TREE	EXPERT
NO	NO	HIGH	REGULAR	HIGH	VENTILATE	VENTILATE
NO	NO	HIGH	REGULAR	REGULAR	VENTILATE	VENTILATE
YES	NO	LOW	REGULAR	HIGH		VENTILATE
NO	NO	LOW	HIGH	LOW		VENTILATE
YES	NO	REGULAR	REGULAR	HIGH		VENTILATE
NO	NO	HIGH	HIGH	LOW	VENTILATE	VENTILATE
NO	NO	LOW	REGULAR	HIGH		VENTILATE
NO	NO	HIGH	REGULAR	LOW	VENTILATE	VENTILATE
YES	NO	REGULAR	HIGH	HIGH		VENTILATE
NO	NO	REGULAR	REGULAR	HIGH	VENTILATE	VENTILATE
NO	NO	HIGH	HIGH	REGULAR	VENTILATE	VENTILATE
NO	NO	REGULAR	HIGH	LOW		VENTILATE
NO	NO	HIGH	REGULAR	HIGH	VENTILATE	VENTILATE
NO	NO	HIGH	HIGH	LOW	VENTILATE	VENTILATE
NO	NO	HIGH	HIGH	REGULAR	VENTILATE	VENTILATE
NO	NO	LOW	REGULAR	REGULAR		VENTILATE
NO	NO	REGULAR	REGULAR	REGULAR	VENTILATE	VENTILATE
NO	NO	HIGH	HIGH	HIGH	VENTILATE	VENTILATE
NO	NO	HIGH	REGULAR	REGULAR	VENTILATE	VENTILATE
NO	NO	REGULAR	HIGH	HIGH	VENTILATE	VENTILATE
NO	NO	REGULAR	HIGH	REGULAR	VENTILATE	VENTILATE
NO	NO	HIGH	REGULAR	HIGH	VENTILATE	VENTILATE
NO	NO	REGULAR	HIGH	REGULAR	VENTILATE	VENTILATE
NO	NO	HIGH	REGULAR	REGULAR	VENTILATE	VENTILATE
TOTAL:					**17**	**24**

pollution and low temperature. In the first aspect the expert decides to open the windows even with outside pollution, while the routing path decides not to if exceeds the standards. In the second aspect (temperature), the expert decides to ventilate also with low temperature even though VOCs do not volatize with low temperature.

5. CONCLUSIONS

The presence of Volatile Organic Compounds inside a non-industrial building represent a health problem on human beings, that it can be regulated with an effective natural ventilation. The Intelligent Environments are a tool to improve the quality of life of the environment of humans in a non-intrusive manner. The EI allows the automation of the opening and closing of the windows in order to maintain an adequate receiving of fresh air of the outside. This will contribute to a suitable Indoor Environment Quality for their occupants. The present paper shows how the Decision Tree is a feasible methodology as a method of learning and machine reasoning. The method allow the construction of a routing path in the taking decision for a computing system that would monitors a real context, based in natural language and real events.The proposed routing path shows how the entropies and uncertainty of the variables involved provide an analytical and suitable method so that can be used in an intelligent system to make the best decision. This routing path represented by the Decision Tree was tested with a PHD expert in Environment Science, from which was obtained a 70.83% of accuracy. It was found that two of the variables that are important in the decision to allow the natural ventilation by the opening of the window are the outside pollution and rain, both represent almost an 86% when adding their entropies.

REFERENCES

1. EPA, US Environmental Protection Agency, «U.S ENVIRONMENTAL PROTECTION AGENCY,» 19 January 2017. [En línea]. Available: www.epa.gov. [Último acceso: 13 October 2018].
2. O. M. d. l. Salud., «Organización mundial de la salud.,» 2018. [En línea]. Available: https://www.who.int/es.
3. Dirección general de industria, energia y minas de la comunidad de Madrid, Guía de Calidad de Aire Interior, Madrid: Arias Montano, 2016.
4. J. M. Sanchez Montero y A. Alcántara León, «Compuestos Orgánicos Volátiles en el medio ambiente».
5. M. Berenger Subils, *NTP 289: Sindrome del Edificio Enfermo.,* España, España: Ministerio de Trabajo y asuntos sociales España.
6. X. Chen, Y. Zheng, Y. Chen, Q. Jin, W. Sun, E. Chang y M. Wei-Ying, «Indoor Air Quality Monitoring System for Smart Buildings,» *UBICOMP,* pp. 471–475, 2014.
7. S. Shaharil Mad, A. Allan Melvin, M. S. Ali Yeon, M. S. Abdul Rahman, A. Muhamad Yusof y A. Zakaria, «Classifying sources influencing Indoor Air Quality (IAQ) using Artificial Neural Network (ANN),» *Sensors,* n° 15, pp. 11665–11684, 2015.
8. F. J. Campuzano Adán, *Metodología para la creación de modelos computacionales de comportamiento humano aplicados a la validación de Entornos Inteligentes,* Murcia: Tesis doctoral, 2015.
9. Secretaria de economia, *NMXA-AA-164-SCFI-2013,* Mexico, 2013.
10. W. H. O. WHO, Guidelines for indoor air quality: selected pollutants, Copenhaguen, Denmark: World Health Organization 2010, 2010.
11. Ministerio de Ambiente y Desarrollo Sostenible, Criterior Ambientales para el diseño y construcción de vivienda urbana, Bogotá: Grupo de Comunicaciones Jose Roberto Arango, 2012.
12. Organización Panamericana de la Salud, «Natural ventilation for infection control in health-care settings,» Organización Panamericana de la Salud, Washington.
13. J. Huamán-Rojas, *Control Inteligente de Sistemas de iluminación en edificios.,* Piura, 2017.
14. D. Archila Córdoba y F. Santamaría Buitrago, «Estado del arte de las Redes de sensores Inálambricos.,» *Tecnología investigacion y Académia,* 2013.
15. E. F. Contreras Morales, F. M. Ferreria Correa y M. A. Valle, «Diseño de un modelo predictivo de fuga de clientes utilizando Arboles de Decisión,» *Revista Ingenieria Industrial,* pp. 7–23, 2017.
16. e. y. m. d. l. c. d. M. Dirección general de industria, «Guía de calidad de aire interior,» Arias Montano, Madrid, España, 2016.
17. E. Gallego Piñol, X. Roca Mussons, M. Rosell Farrás, X. Guardino Sola y E. Gadea Carrera, «Calidad del aire interior: compuestos orgánicos volátiles, olores y confort».
18. A. Muñoz Maña, D. Serrano y F. Sánchez Cid, *Seguridad Dinámica en Ambientes Inteligentes.,* Departamento de Lenguajes y Ciencias de la Computación., 2018.
19. J. Hernández-Aguilar, F. Avila-Camacho, J. Stein-Castillo y A. Mélendez-Ramírez, «Sistema sensor para el monitoreo ambiental basado en redes neuronales.,» *Ingeniería, Investigación y Tecnología,* pp. 211–222, 2016.