Undergraduate Final Project

Occupancy detection in intelligent environments based on low-cost wireless sensor networks and machine learning techniques.

Hyuri da Silva Maciel
hyuri@laccan.ufal.br

Advisor:

Prof. Dr. André Luiz Lins de Aquino

Maceió, 21 de February de 2021

Hyuri da Silva Maciel

Occupancy detection in intelligent environments based on low-cost wireless sensor networks and machine learning techniques.

Monograph presented as a partial requirement to obtain a Bachelor's degree in Computer Science from the Institute of Computing at the Federal University of Alagoas, approved by the examining committee that signs below.

Advisor:

Prof. Dr. André Luiz Lins de Aquino

Maceió, 21 de February de 2021

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

M152o Maciel, Hyuri da Silva.
Occupancy detection in intelligent environments based on low-cost wireless sensor networks and machine learning techniques / Hyuri da Silva Maciel. – 2021.
22 f. : il., figs. e tabs. color.

Orientador: André Luiz Lins de Aquino.
Monografia (Trabalho de Conclusão de Curso em Ciência da Computação) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2021.
Texto em inglês.

Bibliografia: f. 20-22.

1. Redes de sensores sem fio. 2. Aprendizagem de máquina. 3. Ocupação do ambiente (Computação). I. Título.

Monograph presented as a partial requirement to obtain a Bachelor's degree in Computer Science from the Institute of Computing at the Federal University of Alagoas, approved by the examining committee that signs below.

Prof. Dr. André Luiz Lins de Aquino - Advisor
Institute of Computing
Federal University of Alagoas

Leonardo Viana Pereira - Examiner
Institute of Computing
Federal University of Alagoas

Geymerson dos Santos Ramos - Examiner
Institute of Computing
Federal University of Alagoas

Maceió, 21 de February de 2021

Acknowledgment

Now I can say that another cycle of my life has come to an end, where learning went far beyond the academic knowledge, it was where I could find myself in several aspects, mainly as a person. It all started when I was 12 in a small town called União dos Palmares, where for me all the possibilities and perspectives of life were there. But a change in the dynamics of my life happened, and through references and advice from teachers, friends and my mother, I decided to take the exam to attend high school integrated with the technical electronics course at CEFET / IFAL, then I entered the University (UFAL), these two choices were the biggest and best choices I've made in life so far.

My gratitude to my friends, Everton Borges, Matheus Tavares, Tarciso Ferreira, Iago Barboza, João Paulo and everyone who somehow contributed to this journey. To the two professors who changed the trajectory of my life and career at the university, Heitor Ramos and Leonardo Viana who opened the doors to the world of research and took me to LaCCAN.

I want to thank everyone who is part of LaCCAN, Toni, Daise and Thamires for their daily conversations. Demétrios, Geymerson, Randi, Bruno, David, Israel, Matheus, Daniel, Cristopher, Charles, Fabricio, Eduardo, Alvino and all of the “Labs” 15, 16 and 17 for the partnership, conversations and friendship. To all teachers, especially André Aquino (Alla), for believing in me in moments that I did not believed in myself, for the confidence and friendship.

I thank the woman who has the biggest heart in the world and goes to great lengths to help others, who believed in me, did not hesitate to support me, taught me what love is and is my greatest source of inspiration, my mother, thank you very much for your love, i love you so much. To my brother for the reference, to my nieces who make me hope for a better world.

Finally, I thank my friend, companion, partner and love, Caroline, this wonderful woman, who shows me the good side of life, by always supporting me and holding my hand in the most difficult moments, the words written or spoken are too little to express my love for you. I believe that one of the greatest privileges in the world is to love someone who also loves you, I am so lucky.

The journey here was not easy, but all the joy, sadness, insecurity and fear did not let me give up, And none of this would have come true without the references of my life, which have always supported me and never gave up on me, you are mine force.

"Na curva do futuro muito carro capotou talvez por causa disso é que a estrada ali parou. Porém, atrás da curva perigosa eu sei que existe alguma coisa nova mais vibrante e menos triste"..

– Seixas, *R. S.*

Agradecimentos

É, agora posso dizer que chegou ao fim mais um ciclo da minha vida, onde o aprendizado foi muito além do acadêmico, foi onde eu pude me encontrar em vários aspectos, principalmente como pessoa. Onde tudo começou quando tinha 12 anos em uma cidade do interior chamada União dos Palmares, para mim todas as possibilidades e perspectivas de vida estavam lá. Mas uma mudança na dinâmica da minha vida aconteceu, e por referências e conselhos de professores, amigos e minha mãe, decidi fazer a prova para cursar o ensino médio integrado com o curso técnico de eletrônica no CEFET/IFAL, em seguida ingressei na Universidade (UFAL), essas duas escolhas foram as maiores e melhores escolhas que fiz na vida até este momento.

Minha gratidão aos amigos, Everton Borges, Matheus Tavares, Tarciso Ferreira, Iago Barboza, João Paulo e todos que de alguma forma contribuíram para essa jornada. Aos dois professores que mudaram a trajetória da minha vida e carreira na universidade, Heitor Ramos e Leonardo Viana que abriram as portas para o mundo da pesquisa e me levaram para o LaCCAN.

Quero agradecer a todos que fazem parte do LaCCAN, Toni, Daise e Thamires pelas conversas e descontração diárias. Demétrios, Geymerson, Randy, Bruno, David, Israel, Matheus, Daniel, Cristopher, Charles, Fabricio, Eduardo, Alvino e todos dos “Labs” 15, 16 e 17 por toda parceria, conversas e amizade. A todos os professores, em especial André Aquino (Alla), por acreditar em mim nos momentos que eu não acreditava, por toda confiança, amizade.

Agradeço a mulher que tem o maior coração do mundo e não mede esforços para ajudar o próximo, que acreditou em mim, não esitou em me apoiar, me ensinou o que é o amor e é minha maior fonte de inspiração minha mãe, muito obrigado por seu amor, te amo muito. Ao meu irmão pela referência, às minhas sobrinhas que me fazem ter esperança em um mundo melhor.

Por fim agradeço a minha amiga, companheira, parceira e amor, Caroline, essa mulher maravilhosa, que me mostra o lado bom da vida, por sempre me apoiar e segurar minha mão nos momentos mais difíceis, as palavras escritas ou faladas são poucas para expressar o meu amor por você. Acredito que um dos maiores privilégios do mundo é amar alguém que também ama você, essa sorte eu tenho.

A jornada até aqui não foi fácil, mas todas as alegrias, tristezas, inseguranças e medos não deixaram eu desistir. E nada disso teria se tornado realidade sem as referências da minha vida, que sempre me apoiaram e nunca desistiram de mim, vocês são minha força.

"Na curva do futuro muito carro capotou talvez por causa disso é que a estrada ali parou. Porém, atrás da curva perigosa eu sei que existe alguma coisa nova mais vibrante e menos triste".

– Seixas, *R. S.*

Abstract

The smart cities, ideally, can use computer systems who perceive the intention of users, decreasing the need for human intervention to configure the environments. For this, it is necessary to create computational applications that are increasingly adaptable and flexible, improving the services of intelligent environments in a continuous and transparent. In this work, we present a wireless sensor network that detects the occupation in the environment and acts in the lighting and cooling system in an intelligent. Also, some outliers were added to the data in order to validate the results of the classification techniques and check their performance on noise data.

Keywords: Wireless Sensor Network, Machine Learning, Occupation.

Resumo

As cidades inteligentes de forma ideal deveriam utilizar os sistemas computacionais para inferir a intenção do usuários, reduzindo ao máximo a necessidade da intervenção humana para a configuração dos ambientes. Para isto é necessário a criação de aplicações computacionais que sejam cada vez mais adaptáveis e flexíveis, melhorando os serviços dos ambientes inteligentes de forma contínua e transparentes. Neste trabalho apresentamos uma rede de sensores que detecta a ocupação dos ambiente e atua no sistema de iluminação e refrigeração de forma inteligente. Também foram inseridos *outliers* nos dados com intuito de validar as técnicas de classificação em a realização de um pré-processamento nos dados.

Palavras-chave: Rede de Sensores sem fio, Aprendizagem de máquina, ocupação.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

The world population has increased over the years. We expanded from 2.6 billion people in 1950 to 7.8 billion in the year of 2020, and predictions estimate approximately 11 billion by 2100 (**?**). About 90% of people spend most of their time inside their homes (Mohd Nor et al., 2013), and a bigger population will increase energy consumption, which urges us to improve the use of resources such as water and electricity in a world with limited resources.

With the changes in the dynamics of the cities, the population and the government must look for actions that aim at the best use of their resources, either for mobility, communication or energy efficiency. Information and Communication Technologies (ICT) can contribute to these actions by creating solutions that make up smart cities. The six areas that define the central aspects of a smart city are: Smart Economy, Smart Governance, Smart People, Smart Mobility, Smart Environment and Smart Living discussed in (Pellicer et al., 2013).

With the emergence of smart cities, smart environments also appear, which are houses, corporate buildings, industry, and the most varied types of buildings, these environments need a closer interaction with the smart services of cities. They need to provide their preferences for the use of different environments, such as the desired temperature, or have some systems automatically shut down. Smart cities should ideally use computer systems to infer the intention of users, reducing as much as possible the need for human intervention to configure environments settings. Such environments should have the ability to analyze user behavior and act directly, helping individuals. For this, it is necessary to create computational applications that are increasingly adaptable and flexible, improving the services of intelligent environments in a continuous and transparent way.

China is the number one country when it comes to the consumption of electric energy, with a total of 5564 billion (kWh), while Brazil is in the eighth position consuming 509 billion (kWh) in

2020[1]. Although the sudden increase in electricity consumption in large countries, with China being the second-largest world economy and Brazil the ninth, may reflect economic growth and an improvement in the quality of life of the population, it also comes with negative aspects, like the depletion of resources used to generate electrical energy or the impact on the environment, leading to the need for a major investment in research for new sources of energy and energy efficiency solution.

In addition to that, an important fact is that in Brazil the classes that consume the most electricity are industry $34,8\%$, residential $29,6\%$, commercial $19,1\%$, rural $6,0\%$, public power $3,3\%$ and public lighting $3,3\%$. Residential and commercial represent the second and third largest consumers of electricity according to the statistical yearbook for Electric Energy 2020 in 2019, EPE with base [2].

The advance of microelectronics in the creation of sensors with increasing robustness, smaller sizes, and cheaper microcontrollers, push forward the creation of advanced wireless sensor networks (WSNs) for different applications.

Even with better computational power (such as memory and processing), real-time applications require fast responses, and resources must be efficiently used. A wireless sensor network design must consider sensors' battery lifetime, flexibility, data collection, communication protocols, production cost, the environment in which the WSN is installed, and hardware restrictions (Akyildiz et al., 2002).

Wireless sensor network applications for monitoring, forecasting, healthcare, and smart grids (Xu, 2002), present many problems related to the consistency and regularity of the data collected with WSNs (Ganesan et al., 2004). (Corrales et al., 2016) address how to recognize quality issues in regards to heterogeneity, inconsistency, noise, amount of data, etc. Some crucial criteria as the correlation between the collected data must be analyzed to identify outliers. (Zhang et al., 2010) present a discussion on techniques such as statistical-based approaches, the distance between values, and clustering to detect noise in data.

Data pre-processing is crucial for obtaining robust applications since it allows the extraction of the environment's information for characterization. Several problems, such as power outage, battery discharge, and node misplacement, can occur during the acquisition or transmission of data in WSNs.

Even with greater processing power, data analysis with outliers can contribute to delay the system's response, due to the need for cleaning the data. Outliers are values significantly different from the expected values of the data collected by the sensors. Data analysis costs more in these situations due to the difficulty of finding patterns that correctly describe the data.

[1] https://www.indexmundi.com/map/?v=81&l=pt
[2] http://epe.gov.br/

1.2 Related Works

Performing occupancy detection allows the designing of efficient performance policies for smart environments. Previous works of literature report energy consumption reduction between 30 and 42% through building occupancy detection procedures (Erickson et al., 2014). These procedures if applied in addition to machine learning techniques, can reach reduction rates between 29% and 80% (Brooks et al., 2015). Most of a building's energy consumption is directly related to heating and cooling. The growth of HVAC (Heating, Ventilation, and Air Conditioning) systems is significant, and its impact on energy use in European countries represents 76% of total consumption (Nikolaou et al., 2012).

In (Vattapparamban et al., 2016) the tracking of the occupation of people is performed by requesting information from devices with WIFI connection such as smartphones. Its equipment passively collect these signals, which are used to locate users within a defined environment, thus identifying that the zone monitored is busy. The Sentinel system introduced in (Balaji et al., 2013) uses the WIFI network infrastructure of the buildings in conjunction with smartphone devices connected to the network to detect occupation, using this information they were able to save $17,8\%$ of electricity in HVAC systems.

Another approach is presented in (Zou et al., 2018), where it proposes a system for the recognition of human activities, DeeHare, that uses IoT equipment with deep learning application, obtaining $97,6\%$ accuracy in recognition. Passive infrared sensors (PIRs) are widely used to detect occupation in various indoor environments, the use of PIRs sensors do not require a robust infrastructure to be installed, their passive operation makes the PIRs systems last for long periods of time. A wireless sensor network with PIRs sensors is presented in (Lasla et al., 2019) to detect if the environment is busy, a study of the best positioning of the space sensors is also accomplished. The work (Raykov et al., 2016; Wu and Wang, 2018) also demonstrates the use of PIRs sensors to estimate occupancy detection in indoor spaces. The work uses data from temperature sensors of internal and external environments of a school in Spain, the information is collected during daily activities, the data is used to generate an internal temperature prediction model, which can be used to optimize energy consumption (Moreno et al., 2015).

The work in (Han et al., 2012) uses sensors that collect carbon dioxide, relative humidity, and passive infrared connected in a sensor network. The Autoregressive Hidden Markov Model, Hidden Markov and Vector support machine are applied to identify occupancy status, obtaining an average accuracy of $80,78\%$ when using Autoregressive Hidden Markov Model. Using data from environmental sensors, the occupation was determined applying the RF, CART and LDA models, the accuracy achieved was around 98%, considering different sets of pairs of predictors: temperature and light, light and CO2, and light and humidity (Candanedo and Feldheim, 2016).

1.3 Objectives

The main objective of this work is to develop a low-cost wireless sensor network with an easy to implement architecture, able to monitor the environment through sensors and act on equipment through actuators, providing information that can be used for the most varied applications in the context of intelligent buildings. An application is carried out to detect occupation in the environments in which the system is installed, thus performing actions on the lighting and cooling system in order to reduce electricity consumption and make the environment more efficient and comfortable.

1.4 Contributions

In this work, we developed a wireless sensor network with sensors and actuators, it presents an easy to implement architecture, also, we classify the occupation of an environment in the context of smart building applications.

The goal is to reduce the electricity consumption indicators of the monitored environments. We use temperature and luminosity data collected with a WSN. The system also intelligently interferes to control the environment's lighting and cooling devices, with no need for users' interference. Good results of accuracy were obtained with the Random Forest (RF) $98,48\%$, Classification-and-Regression-Tree (CART) $96,56\%$, and the K-Nearest Neighbors (KNN) $98,40\%$ algorithms. Inserting 10% of outliers in the data and applying the RF, CART and KNN techniques, the accuracy values are $97,18\%$, $95,00\%$ and $97,04\%$ respectively.

Due to the pattern similarity of the collected data, we added some disturbances in the collection and transmission processes. Homogeneous patterns are not common in uncontrolled environments, where the occupants' behavior may suddenly change. We evaluate the performance of the classification techniques subjected to the noisy data created with the disturbances, and presented a context-aware application, using sensors to detect occupancy in buildings with classic classification techniques. The inserted noise simulates sensor failure and atypical behavior in the environment.

1.5 Structure of Text

The text structure follows as Chapter 2 presenting the technologies and concepts involved in the work presented, in Chapter 3 we present our proposal, Chapter 4 presents the results and finally, Chapter 5 contains the conclusion and final considerations.

2

Methodologies and Concepts

This section details the methodology by describing the techniques and technologies used in this work. In our results, using data from environmental temperature and humidity sensors, we determined the occupation of the environment using RF, CART and KNN classification techniques. The evaluation showed that this is a viable approach to detect occupancy in environments and to reduce energy consumption. Next, we will continue with the discussion of the configured RSSF and the techniques applied to the collected data.

2.0.1 Wireless Sensor Network

Wireless sensor networks (WSN) are sensor nodes coupled with a processor, a communication interface, sensors, and a power supply (Rawat et al., 2014). Some sensor nodes may have an internal memory and analog-to-digital converters. WSNs can have thousands of sensor nodes or units, have the ability to monitor various types of environmental conditions, such as temperature, humidity and in some cases act in the environment by activating a lamp or opening a door, and must have self-configuration mechanisms due to failures of communication and problems in the nodes.

Many factors directly influence the design of a WSN, some of the main ones are: fault tolerance, scalability, production costs, operating environment, network topology, hardware restrictions, transmission media and energy consumption (Akyildiz et al., 2002).

Our WSN has two types of nodes, a central node called SINK that collects data from the other sensors in the network, who are responsible for collecting data from the environment and acting on their devices, these two types of nodes have limited memory and processing power, but they still allow the execution of machine learning techniques, thus performing a faster response. The WSN uses a Radio communication interface, its source of energy is the electrical energy of the environment, connected directly to the building's power grid, making it not necessary to change batteries or recharge the nodes, resulting in a continuous data collection while the environment has electricity.

WSN communication uses a Mesh architecture, which allows any node connected to the network to transmit to any other node that is also within its transmission range, also known as multihop communication (Townsend and Arms, 2005). If the node that is transmitting data has no reach with the sink node it can use an intermediate node to transmit the data, this type of network also has the characteristic of auto configuration, in case a node fails.

The application's WSN collects the following data: temperature, humidity and luminosity. The environment's electric current data can also be monitored.

The system's prototype has three nodes: `Sink`, `node_01`, and `node_02`. It was designed considering low-cost requirements, environment data collection, and actuation upon devices such as lamps, windows, blinds, etc. By inserting other sensors, the prototype can be extended or adapted to other applications and different environments.

The `node_sink` Figure 2.1 is composed of a Raspberry PI 3 Model B+ [1] (device `1`), with Broadcom BCM2837B0 64-bit ARM Cortex-A53 Quad-Core processor, clocked at 1.4GHz, 1GB RAM, WiFi 802.11 b/g/AC 2.4GHz, connected to a NRF24L01 WIFI module (device `2`) (Sem, 2008), an electronic component that sends and receives data via wireless and can be integrated with several microcontrollers, like Arduinos, ESP, Raspberry among others, with a range up to 1 km away in outdoor environments, transmitting data at a programmable rate of 250kbps to 2MB, with an operating frequency of 2400MHz to 2524MHz, This node receives data collected through the WSN, executes inference algorithms, and disseminates actuation commands to other sensor nodes in the network.

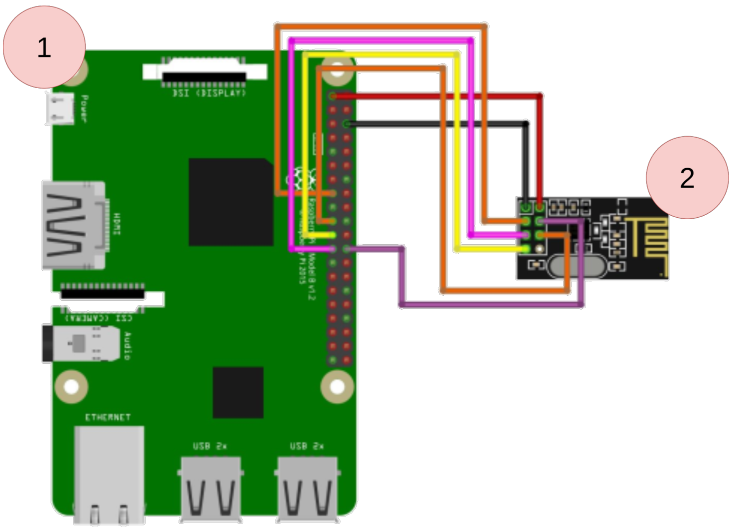

Figure 2.1: Schematic representation of the central node (SINK) of the WSN.

[1] https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/

`Node_01` 2.2 is composed of an Arduino Uno [2] (device `1`), which is an open-hardware prototyping platform, and a radio module NRF24L01 (device `2`) for communication.

Features of the Arduino UNO:

- Microcontroller ATmega328;
- Input Voltage $7-12V$;
- Digital I/O Ports: 8, plus 6 (Digital or PWM);
- Analog Ports: 6;
- I/O Pin Current: $40mA$;
- Flash memory: $32KB$;
- SRAM: $2KB$;
- EEPROM: $1KB$;
- Clock Speed: $16MH_z$.

Also the `Node_01`, has a DHT11 sensor for monitoring temperature and humidity (device `3`) (Gay, 2018), a BH1750 light sensor (device `4`) (ROH, 2010) which is composed of a 16-bit internal digital analog converter and communicates via I2C with the microcontroller, a Infrared IR 5mm transmitter (device `5`) (Vis, 2014), that sends signals mimicking the air conditioning's remote control as programmed in the microcontroller with an operating voltage of $1.2V-1.4V$ and a wave-length of $940nm$, and four relays (devices `6`) to make it possible to activate devices of 220V AC, such as lamps, electronic equipment, motors, and also to isolate one circuit from another, acting as magnetic switches to turn the lamps on and off.

Features of the sensor DHT11:

- Humidity measuring range: $20-90\%$ UR;
- Temperature measurement range: 0^o-50^oC;
- Input Voltage: $3-5VDC$;
- Current: $200\mu A-500mA$;
- Humidity measuring accuracy: $\pm\,5,0\%UR$;
- Temperature measurement accuracy: $\pm\,2.0^oC$.

Features of the sensor BH1750:

[2] https://store.arduino.cc/usa/arduino-uno-rev3

- Light sensor: GY-302;
- Input Voltage: $3-5VDC$;
- Measuring range: $1-65.535Lux$;
- Interface communication: I2C;
- Resolution: $1Lux$.

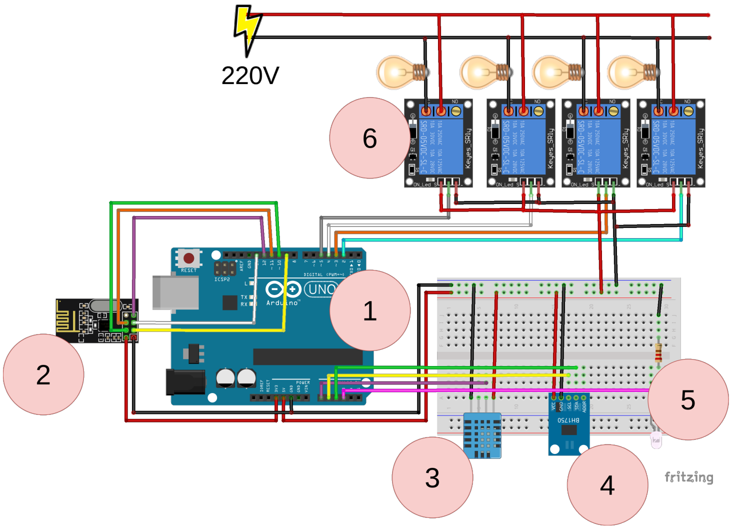

Figure 2.2: schematic representation of the WSN `node_1`, consisting of sensors and actuators.

The equipment that gathers data on electric energy consumption `node_02` is represented by Figure 2.3. This device is composed of an Arduino UNO (device `1`), a NRF24L01 module (device `2`), current sensors, SCR013 (devices `3`) used to measure alternating current (AC) of up to 20A in a non-invasive way, and can be used to monitor equipment such as motors, computers and the power generation board of the building or monitored environment, and an RTC (*Real Time Clock*) (device `4`) (Max, 2015) with 56 bytes of SRAM that is able to inform and save the current time, relying on its internal battery that guarantees the clock operation even without power.

Features of the sensor SCR013:

- Model sensor: SCT-013-020;
- Input current: $0-20A$;
- Output signal: $1V$;
- Core Material: Ferrite.

Features of the sensor RTC:

- Chip: DS1307;

- Interface: I2C;
- Power failure detection circuit.

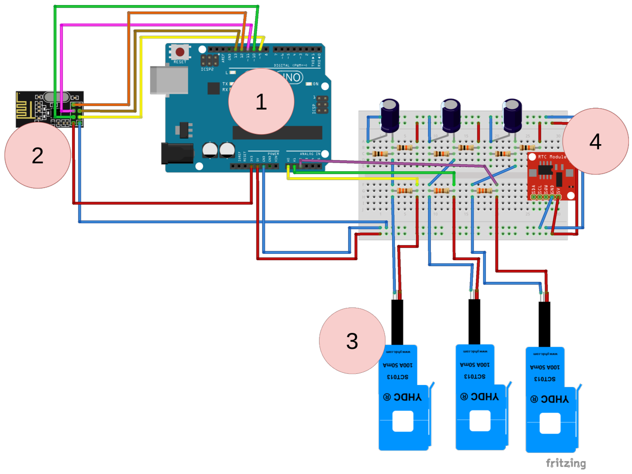

Figure 2.3: schematic representation of the WSN `node_2`, responsible for monitoring the electrical consumption of the equipment.

The ABNT NBR 5413 (Ass, 1992) standard governs the rules that determine the illuminance values for artificial lighting, in the most diverse environments, ensuring adequate lighting conditions for the most varied activities, the characteristics of the task are observed, the age group of the users is observed, the speed at which the tasks must be performed, and the background brightness, for each observation, weights are assigned varying from -1, 0 or $+1$, the sum of the weights determines the appropriate illuminance group. When the value is -2 or -3 a low illuminance is set to the environment, when it is $+2$ or $+3$ a higher illuminance is set, for the other values a average lighting setting is recommended.

Following the norm NBR 5413 we can make the most of the external natural lighting, and configure the lamps to perform the compensation during the day, also ensuring a comfortable environment for carrying out the activities. For example, in the environment where the data was collected, the age of the users was less than 40 years old which was assigned with a weight equal to -1, speed and accuracy of the performed activity in the environment and background reflectance of the task (between 30 to 70%) were considered important features, and therefore were both assigned to a weight 0, obtaining an algebraic sum of -1, then it is recommended to use the average illuminance value.

2.0.2 Occupation detection

The Figure 2.4 represents the environment where the sensors were placed, about 6 people attend the space on a daily basis. It is worth mentioning that the system can be installed in other environments. The deployment of the sensors is an important WSN aspect, as it impacts directly the gathered data. For example, the temperature collected near windows or air conditioning has different variations. In order to obtain consistent results, we placed the sensor nodes in specific locations less subject to light and temperature variations.

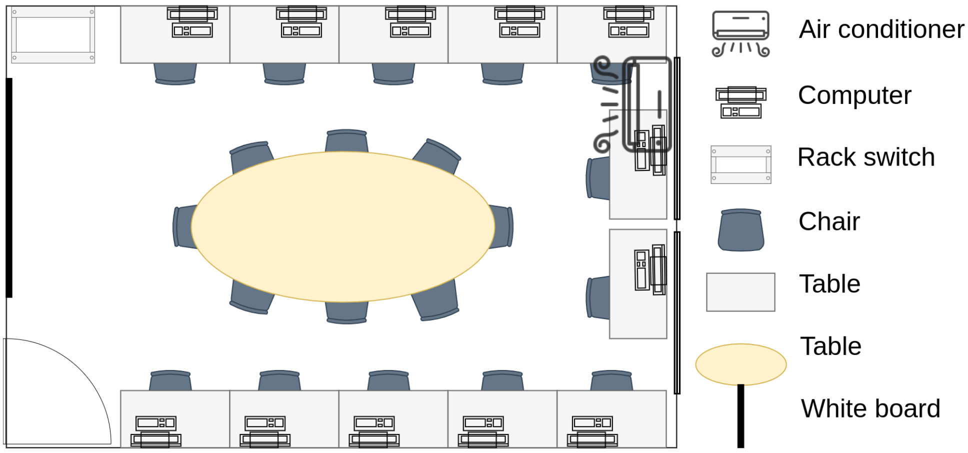

Figure 2.4: Environment where the sensory data and occupation status were collected, has the dimensions of 55x90.

A video camera was configured to record the room and to help us create a labeled data set. We establish the occupation state $s = 1$, if the room is occupied by at least one person, otherwise $s = 0$. The data set is validated through human inspection and used by this work's machine learning classification algorithms. Since the environment's occupants are usually the same people with the same daily routines, it is possible to identify data patterns in temperature and light measurements. Table 2.1 presents a data sample of temperature and light. The ID field represents the unique identifier of the environment. There is also temperature, relative humidity, luminosity, date and time, and the environment occupancy status.

Table 2.1: Sample of the data collect with the sensors.

ID	Temperature	Humidity	Luminosity	Date	Occupancy
6	22.80	54.69	40.56	$2018-10-20\ 17:07:42$	0
6	22.84	54.35	40.46	$2018-10-20\ 17:08:12$	0
6	33.67	54.02	38.97	$2018-10-20\ 17:08:42$	0
6	22.82	54.27	38.87	$2018-10-20\ 17:09:12$	0

Due to the binary nature of the occupation problem, we chose the Random Forest (RF), Classification and Regression Tree (CART), and the K-Nearest Neighbors (k-NN) techniques. Countless literature works have shown that these algorithms can perform decently well in classification problems (Candanedo and Feldheim, 2016).

Classification Techniques

The Random Forest (RF) uses a supervised learning approach, in (Breiman, 1996) where randomness is added, building each tree using a *bagging* sample, each node of the random forest is divided observing the best among the subset of prediction variables chosen randomly in the node, it is robust and it avoids over-fitting. It can be used in large volumes of data, such as those collected by WSNs.

By applying the RF algorithm, we can avoid the problem of data with too many dimensions. From a simplified view, RF randomly chooses n training samples from the data set and creates multiple decision trees, each one generating an output. The final result is achieved by majority voting or averaging of the individual trees' results.

The Classification and Regression Tree Algorithm (CART) generates binary trees by repeatedly splitting one node into two child nodes. CART defines a set of rules or thresholds that divides each child node in the tree and leads to the terminal nodes that are the classes used as prediction results (Loh, 2011).

The K-Nearest Neighbor Algorithm (K-NN) can be applied to classification problems, in which data is grouped based on its similarity value (Lantz, 2015). KNN requires applying a similarity (or distance) function between two points. Manhattan, Minkowski, and Euclidean distance 2.1 are commonly used as functions to measure the distance between two instances. The k-elements that are most similar receive the same label, and the value of k can impact how well the model can generalize predictions on future input.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \ldots + (x_n - y_n)^2} \tag{2.1}$$

Noise Generation

In this work, Bernoulli distribution represents sensor failures and different environment dynamics. We used the Bernoulli distribution to determine which value of the collected data set would be replaced by noisy data. The replacement happens according the probability $p \in [0,1]$, following the Bernoulli distribution and its random variable $x \in \{0,1\}$, such that $\mathbb{P}(X = 0) = 1 - p$ and $\mathbb{P}(X = 1) = p$ (Wasserman, 2013). The Bernoulli probability function is given by Equation (2.2).

$$f(x) = p^x (1-p)^{x-1} \tag{2.2}$$

$$x \in \{0,1\} \tag{2.3}$$

A Bernoulli series was generated with the same dimension as the database collected by the WSN, when a Bernoulli probability was $P = 1$ an outlier was generated with the value established in the intervals ($0 \leq T_n \leq 65,555$ for the lighting sensor and $0 \leq T_n \leq 24.92$ and

$28.92 \leq T_n \leq 39.00$ for the temperature sensor) using the runif [3] function of R, the value generated was replaced by the original value that was in the WSN database and the same position as the result of the Bernoulli serie.

The values of outliers were chosen observing the maximum and minimum values of the sensors and their quartil. For the temperature sensor, we used the lowest value 22.02 until the first quartil 24.42 and from the third quartil 28.84 until the maximum value collected 29.67, which would represent a different dynamic of the use of the room, and the interval 0 to 22.02 which represents a failure in the sensor, because in the geographic region where the laboratory is located, low temperatures are rare, setting the range represented from $0 \leq T_n \leq 24.92$ and $28.92 \leq T_n \leq 39.00$. And for the lighting sensor the range was from $0 \leq T_n \leq 65,555$, because in the environment the maximum value collected was 99.24 Lux, even at night when there was no occupation in the environment, this is due to the external lighting that comes through windows, so we chose the range up to 65,555 as it is the highest Lux value that the sensor can collect representing different dynamics and sensor failures.

[3] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Uniform

3

Our proposal

3.1 Prototype

Figure 3.1 illustrates a possible the WSN's nodes placement in the environment, in which `node_01` is close to the lamps' switches, and each switch controls four lamps numbered 1, 2, 3, and 4 as shown in the figure. The temperature, light, and humidity sensors are approximately 1.5 meters above the floor, far from windows and the air conditioner. The sink node is placed arbitrarily on a workstation (for convenience), while `node_02` is collecting electric current data from the air conditioner. We could also use `node_02` to collect data from devices such as computers and the room's video projector.

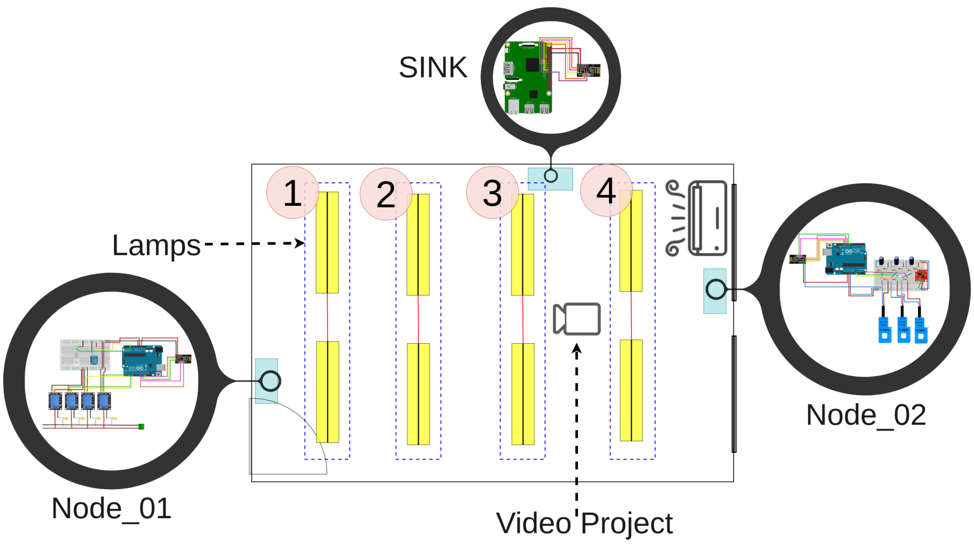

Figure 3.1: WSN's nodes placement: node_01 controls lighting and cooling circuits, node_02 collects air conditioner's electric current data, Sink performs inference operations.

Table 3.1: Cost of the WSN implemented from LaCCAN.

Item	Amount	Value Un. ($R\$$)	Value total ($R\$$)
Raspberry	1	$349,00	$349,00
Arduino Uno	2	$51,11	$102,22
RF24L01	3	$39,90	$119,70
Relay	4	$9,90	$39,60
DHT11	1	$13,90	$13,90
BH1750	1	$18,90	$18,90
Resistor 10k ohms	6	$00,05	$00,30
Resistor 330 ohms	4	$00,05	$00,20
Capacitor	3	$00,21	$00,63
SCR013	3	$54,90	$164,60
RTC clock	1	$18,90	$18,90
Total	22	$556,82	$827,95

3.2 System architecture

In this work we also propose an architecture represented in the figure 3.2 for the implementation of the system, where the main characteristic is that it is low cost and capable of integrating with the most diverse types of sensors and actuators. Having four main layers, environment, sensor devices, control devices and application.

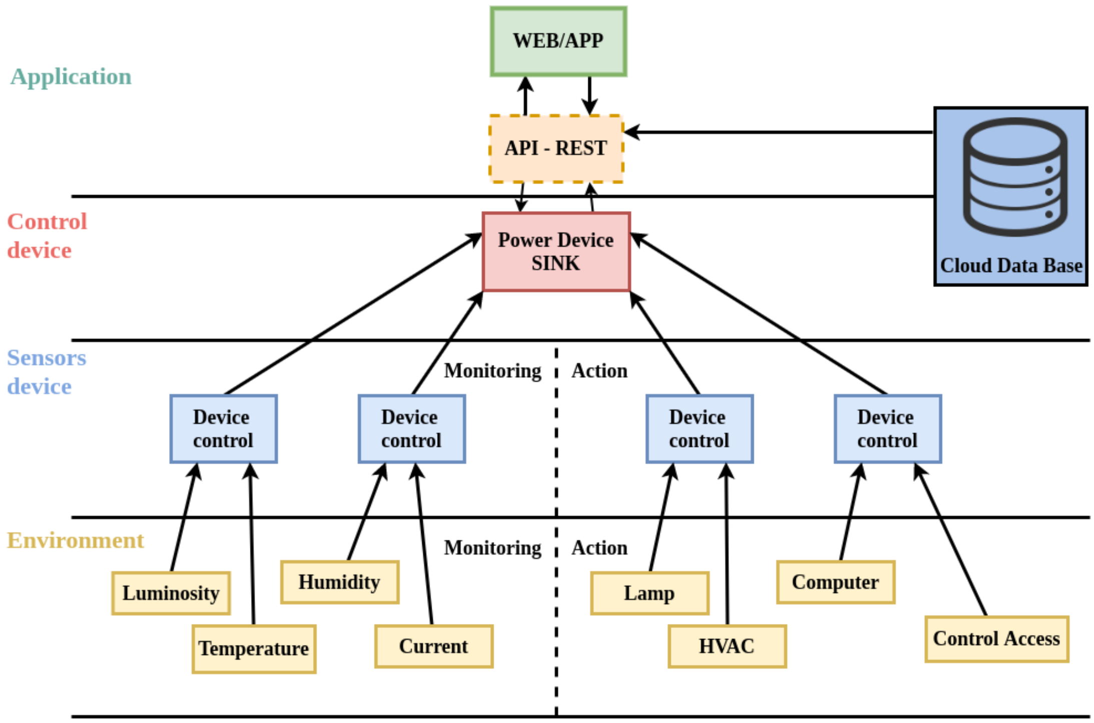

Figure 3.2: Temperature, humidity, light data, and the occupation status.

The first layer, environment, represents the spaces where the sensors and actuators are inserted, the sensors are of the most varied types such as pressure, light, gas sensors, etc. They can have several communication interfaces such as I2C, UART, USB, among others, the actuators are motors, magnetic switches, any electronic or electromechanical device capable of acting on the furniture, electronics and peripherals in the environment.

Sensor devices represent the second layer, they are the microcontrollers where the sensors and the communication module are integrated, they can be represented by Arduinos, ESP, Beaglebone, MICAz. This layer may have the ability to make decisions about local actions and perform small initial treatments on the data.

Control devices, represents the central node layer of the WSN generally called SINK, this layer should be the node with the most processing power in the WSN, because this is where the decision making is carried out to act on the devices of the environment, in this layer we must ensure that the node has the necessary capacity to apply the algorithms, receive and transmit the information to the network without generating delays or problems, either with communication or network operation.

Finally, the application layer, where the data is sent to servers in the cloud, preferably FOGs, must be stored and used when necessary to train a new technique or be accessed by its users through a web application. The exchange of information must be carried out by a REST API, ensuring quick access to the data. It is worth mentioning that this layer is not the main scope of this work.

3.3 Application

With the objective of reducing costs and using the environment in which the system is inserted more efficiently, among the most varied possible applications for the data collected through the WSN, this work focuses on energy savings, identifying space occupation. With that being able to make decisions like turning electronic equipment on or off, for that we apply machine learning techniques, more specifically classification techniques. In order to validate and identify how the result of the classification techniques behave with problems in the WSN, noises were inserted to simulate failures in the sensors and in the collected data. These results are discussed in the chapter 4.

4

Results

In this section, we will present the results for the occupation detection with RF, KNN, and CART algorithms. Figure 4.1 represents charts for temperature, humidity, luminosity and the state of occupation of the environment on August 21, 2018.

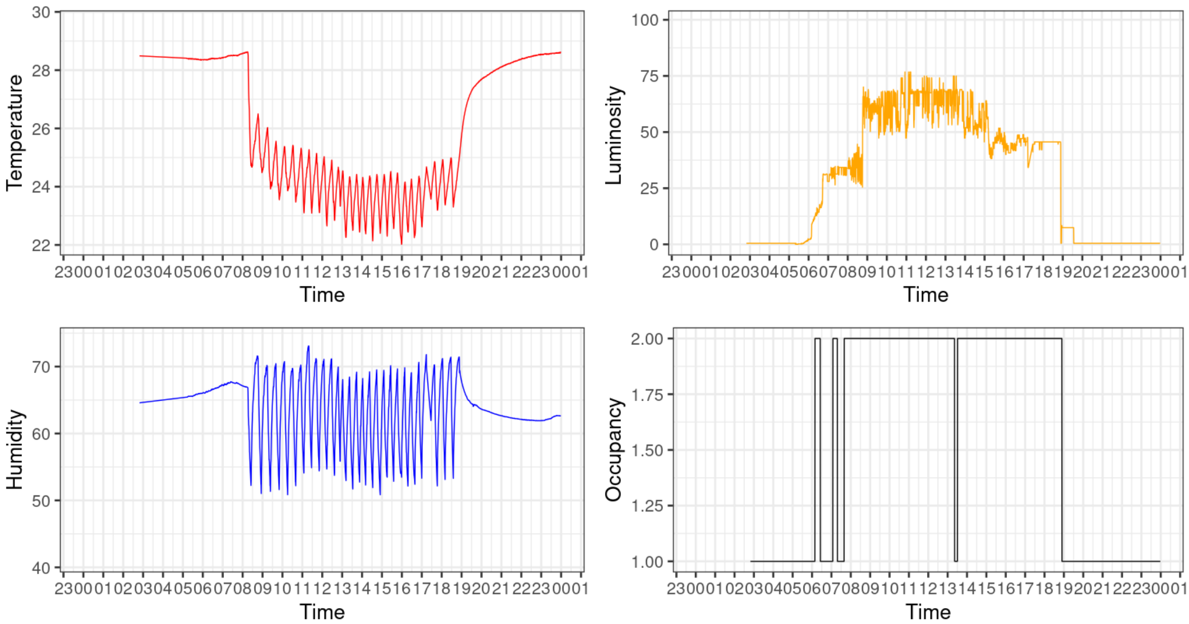

Figure 4.1: Temperature, humidity, light data, and the occupation status.

During work hours, from 00:01 AM to 23:59 PM (00:01 to 23:59 in Figure 4.1), as people arrive and the air conditioner is on, we can observe a temperature decrease. Luminosity increases as the lamps are turned on, there is also sunlight influence until 06:00 PM (18:00). Humidity changes while the air conditioner is on. Usually, the monitored room is occupied between 9:00 AM and 07:00 PM. Mostly, this pattern can be seen every weekday, changing significantly during weekends and holidays. Table 4.2 presents the RF and CART algorithm's classification results, considering temperature and light predictors. While the CART algorithm was able to achieve a

96.56% accuracy in the test data set, better results are provided by the RF technique, which achieved 98.48% accuracy.

Table 4.1: The RF and CART classification results for the temperature (T) and light (L) predictors.

Models	Predictors	Train accuracy	Test accuracy
RF	T and L	0.9965	0.9848
CART	T and L	0.9665	0.9656

In regards to the k-NN algorithm, which is validated through cross-validation, the best performance was achieved when $k = 2$, providing a 98.40% accuracy, very close to the RF algorithm. At first, we can conclude that their results are statistically equal, but we will see that the k-NN has better results when subjected to noisy data samples.

Figure 4.2 and 4.3 presents a example of temperature and luminosity samples composed by 1% and 3% noisy values respectively, which occur according to the Bernoulli variable $x = 1$.

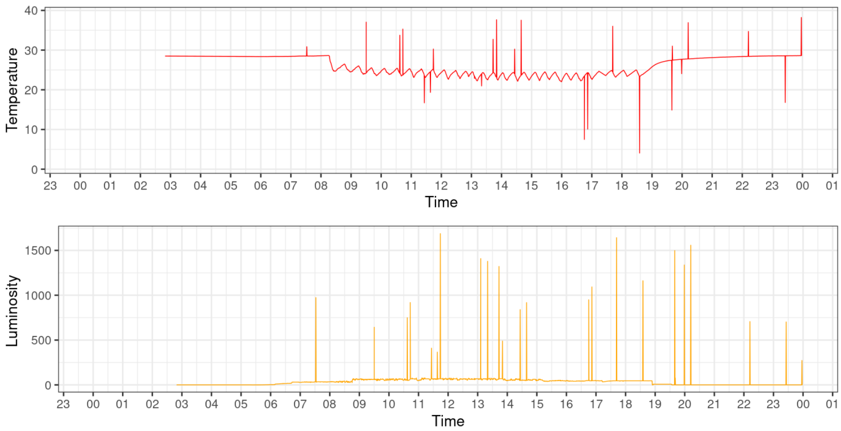

Figure 4.2: Temperature, humidity, light data, and the occupation status.

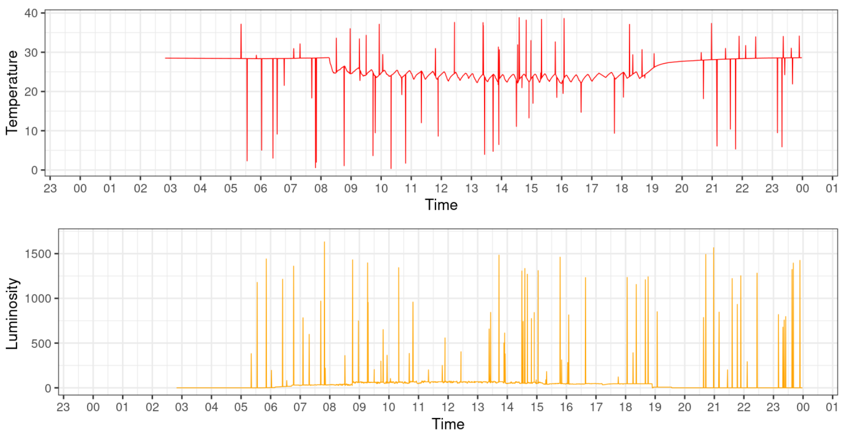

Figure 4.3: Temperature, humidity, light data, and the occupation status.

Following the Bernoulli probability function to verify the algorithm's performance under abnormal sensor readings, we considered five noise rate cases: 1%, 2%, 3%, 5% and 10%. We apply the RF and CART techniques to temperature and luminosity predictors. As shown in Table 4.2, noting that the RF algorithm should decrease its accuracy by just over 1% even when the data has 10% noise, the same can be observed for CART technique.

Table 4.2: RF and CART accuracy evaluation in noisy data samples.

Technique	Noise Rate	Train accuracy	Test accuracy
RF	1%	0.9965	0.9781
CART	1%	0.9586	0.9611
RF	2%	0.9965	0.9737
CART	2%	0.9510	0.9585
RF	3%	0.9961	0.9722
CART	3%	0.9461	0.9526
RF	5%	0.9967	0.9774
CART	5%	0.9463	0.9552
RF	10%	0.9974	0.9718
CART	10%	0.9552	0.9500

The k-NN's noisy data evaluation, presented in Table 4.3, achieved better results for $K = 2$. In this case, its best performance was for noise rates between 1 % and 3%. These results, as similarly argued to the RF and CART algorithms, show that the k-NN is also a good solution to detect occupation.

Table 4.3: k-NN's evaluation results for the noisy data samples, $k = 2$.

Noise Rate	**Accuracy**
1%	0.9807
2%	0.9759
3%	0.9733
5%	0.9770
10%	0.9704

5

Final Remarks

A low cost network of sensors with excellent processing power was built, that can be integrated with other types of sensors and actuators in a simple way. Based on the premise of a robust, low-cost and fully scalable system, we developed an architecture for building the system in other contexts, where this work developed applications up until the third layer of the architecture.

An application to reduce electrical consumption by detecting occupancy of the environment was performed using classification techniques. Using only two predictors temperature and lighting, the Random Foreste, CART and KNN techniques performed well for this problem, where RF achieved a better performance with an accuracy of 98.48%.

As our data has a characteristic routine, which is not reality in several corporate environments, for example, our sensors also did not present major flaws in the observed period, we inserted outliers to represent other types of behavior and flaws in the WSN in order to verify the performance of the techniques used, noise was inserted in 1%, 2%, 3%, 5% and 10% of the data. We can observe that with 10% of the data, an accuracy of 97.18%, 95.00% and 97.04% was obtained for the RF, CART and KNN techniques respectively, showing that even with noise the techniques showed excellent performances, since 10% errors in the data is not usual for a WSN application.

For future work we intend to insert the data of the electric consumption sensors to check how much we have achieved in reducing energy consumption, with the application of the RF technique and without it. Also we intend on applying new inference techniques in the environment such as thermal comfort, thereby making the environment increasingly sensitive to context.

References

Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

Iluminância de interiores NBR 5413. Associação Brasileira de Normas Técnicas, 1992.

Bharathan Balaji, Jian Xu, Anthony Nwokafor, Rajesh Gupta, and Yuvraj Agarwal. Sentinel: occupancy based hvac actuation using existing wifi infrastructure within commercial buildings. volume 17, pages 1–14. Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, 2013.

Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

Jonathan Brooks, Saket Kumar, Siddharth Goyal, Rahul Subramany, and Prabir Barooah. Energy-efficient control of under-actuated hvac zones in commercial buildings. *Energy and Buildings*, 93:160–168, 2015.

Luis M Candanedo and Véronique Feldheim. Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. *Energy and Buildings*, 112:28–39, 2016.

David Camilo Corrales, Agapito Ledezma, and Juan Carlos Corrales. A systematic review of data quality issues in knowledge discovery tasks. *Revista Ingenierías Universidad de Medellín*, 15(28):125–150, 2016.

Varick L Erickson, Miguel Á Carreira-Perpiñán, and Alberto E Cerpa. Occupancy modeling and prediction for building energy management. *ACM Transactions on Sensor Networks*, 10(3): 1–28, 2014.

Deepak Ganesan, Sylvia Ratnasamy, Hanbiao Wang, and Deborah Estrin. Coping with irregular spatio-temporal sampling in sensor networks. *ACM SIGCOMM Computer Communication Review*, 34(1):125–130, 2004.

Warren Gay. Dht11 sensor. In *Advanced Raspberry Pi*, pages 399–418. Springer, 2018.

Zhenyu Han, Robert X Gao, and Zhaoyan Fan. Occupancy and indoor environment quality sensing for smart buildings. pages 882–887, 2012.

Brett Lantz. *Machine Learning with R*. Packt Publishing Ltd, second edition, jul 2015.

Noureddine Lasla, Messaoud Doudou, Djamel Djenouri, Abdelraouf Ouadjaout, and Cherif Zizoua. Wireless energy efficient occupancy-monitoring system for smart buildings. *Pervasive and Mobile Computing*, 59:101037, 2019.

Wei-Yin Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):14–23, 2011.

DS130764 x 8, Serial, I2C Real-Time Clock. Maxim Integrated Products, 2015.

Nursyarizal Mohd Nor, I Elamvazuthi, Nallagownden Perumal, and Ibrahim Taib. Robust stochastic control model for energy and comfort management of buildings. *Australian Journal of Basic and Applied Sciences*, pages 137–144, 2013.

M Victoria Moreno, Antonio F Skarmeta, Alberto Venturi, Mischa Schmidt, and Anett Schuelke. Context sensitive indoor temperature forecast for energy efficient operation of smart buildings. pages 705–710, 2015.

Triantafyllia G. Nikolaou, Dionysia S. Kolokotsa, George S. Stavrakakis, and Ioannis D. Skias. On the application of clustering techniques for office buildings' energy and thermal comfort classification. *IEEE Transactions on Smart Grid*, 3(4):2196 – 2210, dec 2012.

Soledad Pellicer, Guadalupe Santa, Andres L Bleda, Rafael Maestre, Antonio J Jara, and Antonio Gomez Skarmeta. A global perspective of smart cities: A survey. pages 439–444, 2013.

Priyanka Rawat, Kamal Deep Singh, Hakima Chaouchi, and Jean Marie Bonnin. Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of supercomputing*, 68(1):1–48, 2014.

Yordan P Raykov, Emre Ozer, Ganesh Dasika, Alexis Boukouvalas, and Max A Little. Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, volume 16, pages 1016–1027, 2016.

Digital 16bit Serial Output type ambient light sensor IC. ROHM Co., Ltd., 2010. 2010.4.

nRFL24L01+ prelimirary Product Specification. Semiconductor, Nordic, 2008. 1.0.

Chris Townsend and Steven Arms. Wireless sensor networks. *MicroStrain, Inc*, 20(9):15–21, 2005.

Edwin Vattapparamban, Bekir Sait Çiftler, Ismail Güvenç, Kemal Akkaya, and Abdullah Kadri. Indoor occupancy tracking in smart buildings using passive sniffing of probe requests. pages 38–44, 2016.

High Power Infrared Emitting Diode, 940 nm, GaAlAs, MQW. Vishay Semiconductors, 2014. 2.4.

Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

Libo Wu and Ya Wang. A low-power electric-mechanical driving approach for true occupancy detection using a shuttered passive infrared sensor. *IEEE Sensors Journal*, 19(1):47–57, 2018.

Ning Xu. A survey of sensor network applications. *IEEE communications magazine*, 40(8): 102–114, 2002.

Yang Zhang, Nirvana Meratnia, and Paul Havinga. Outlier detection techniques for wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, 12(2):159–170, 2010.

Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J Spanos. Towards occupant activity driven smart buildings via wifi-enabled iot devices and deep learning. *Energy and Buildings*, 177: 12–22, 2018.