Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels
UNIVERSITAT POLITÈCNICA DE CATALUNYA

MASTER THESIS

TITLE: ANALYSIS AND EVALUATION OF WI-FI INDOOR POSITIONING SYSTEMS USING SMARTPHONES

MASTER DEGREE: Master in Science in Telecommunication Engineering & Management

AUTHORS: David Hinojosa Muñoz

DIRECTOR: Roc Meseguer Pallarès

DATE: November, 4th 2016

Títol: ANALYSIS AND EVALUATION OF WI-FI INDOOR POSITIONING SYSTEMS USING SMARTPHONES

Autor: David Hinojosa Muñoz

Director: Roc Meseguer Pallarès

Data: 4 de novembre del 2016

Resum

Aquest document tracta d'analitzar els principals algoritmes utilitzats en Machine Learning aplicats a la localització d'interiors. Les noves tecnologies s’enfronten a nous reptes. El posicionament per satèl·lit s'ha convertit en una aplicació habitual dels telèfons mòbils, però deixa de funcionar satisfactòriament en espais tancats. Actualment hi ha un problema en el posicionament que està sense resoldre. Aquesta circumstancia motiva la investigació de nous mètodes.

Després de la introducció, el primer capítol exposa els mètodes actuals de posicionament i planteja el problema de la localització d'interiors. Aquesta part del treball mostra de manera global l'estat actual de l'art. Esmenta una taxonomia que ajuda a classificar els diferents tipus de posicionament d'interiors i es mostra una selecció actual de solucions comercials.

El segon capítol és molt més focalitzat en els algoritmes que seran analitzats. S'explica el funcionament matemàtic dels algoritmes més usats en Machine Learning. L'objectiu d'aquesta secció és donar a conèixer de manera teòrica algoritmes matemàtics. Aquests algoritmes no van ser dissenyats per a la localització d'interiors però poden ser usats per infinitud de solucions.

En el tercer capítol, ens fa conèixer les eines de Treball: Weka i Python. Es mostren els resultats obtinguts després de milers d’execucions amb diferents paràmetres i algoritmes que mostren els principals problemes de Machine Learning.

En el quart capítol s'analitzen els resultats i es mostren les conclusions obtingudes després analitzar diferents problemes de la localització d'interiors i els diferents algoritmes de Machine Learning.

Title: ANALYSIS AND EVALUATION OF WI-FI INDOOR POSITIONING SYSTEMS USING SMARTPHONES

Author: David Hinojosa Muñoz

Director: Roc Meseguer Pallarès

Date: September 23 rd 2016

Overview

This paper attempts to analyze the main algorithms used in Machine Learning applied to the indoor location. New technologies are facing new challenges. Satellite positioning has become a typical application of mobile phones, but stops working satisfactorily in enclosed spaces. Currently there is a problem in positioning which is unresolved. This circumstance motivates the research of new methods.

After the introduction, the first chapter presents current methods of positioning and the problem of positioning indoors. This part of the work shows globally the current state of the art. It mentions a taxonomy that helps classify the different types of indoor positioning and a selection of current commercial solutions.

The second chapter is more focused on the algorithms that will be analyzed. It explains how the most widely used of Machine Learning algorithms work. The aim of this section is to present mathematical algorithms theoretically. These algorithms were not designed for indoor location but can be used for countless solutions.

In the third chapter, we learn gives tools work: Weka and Python. the results obtained after thousands of executions with different algorithms and parameters showing main problems of Machine Learning shown.

In the fourth chapter the results are collected and the conclusions drawn are shown.

Acknowledgements

In these first lines of the Thesis, I take the opportunity to especially thank my Tutor Roc Meseguer, for having guided the selection of my Project Master Thesis and for his valuable assistance in the development thereof. I also want to thank him for always being available when needed. He has been very patient and encouraged me to continue in the moments that most needed.

Given my work situation, this project has been extended in time more than is desirable. I appreciate the support of my family and my beloved. They were always with me, and despite my absence at many times, were sympathetic. They helped me get up after each setback and understood that I could not be. I appreciate the faith they had in me.

ÍNDEX

INTRODUCTION

The maps were prized by kings and sages. They gave the key of the kingdoms and the possibility of overcoming battles. When we decided to explore the world, positioning was a matter of life or death. To know where we are, we need points of reference. In the past we used the stars. The sky was guiding us. Over time we built our own reference points on earth to guide us into the sea, the headlights. Initially guided by light, but then, they guided by radio frequency. Shortly after, our warlike nature forced us to sophisticate technology. We put our own stars in the sky. Lighthouses located hundreds and thousands of miles beyond the clouds giving global coverage. Despite the military nature of satellite guidance systems currently used in many commercial solutions. The evolution of computing and digital signal processing has enabled miniaturize the reception enabled devices. In 2015, there were 1685 billion smartphones in the world. All of them with the possibility of geolocation. We tend to a global world, where geolocation has a significant presence and there are still unresolved challenges.

The positioning by GNSS resolved positioning outdoors. GNSS work by counting the time of arrival of the signal from satellite to the device. If we know the speed of propagation of the signal, we can calculate the distance between them. As the satellite position is known at all times, the terminal is able to know where it is, as long as there are visible three satellites or more. The distance between the satellite and the terminal describes a radius of a circumference. The three circumferences meet at a point. That point is the position of the terminal. This method requires the terminal and the satellite have a direct view. Otherwise, the calculation of distance fail. Indoors, the signal is much attenuated and suffers rebounds. Attenuation can make the signal should not be perceived by the terminal. The effect multipath delays the arrival of the signal and causes an error in the location.

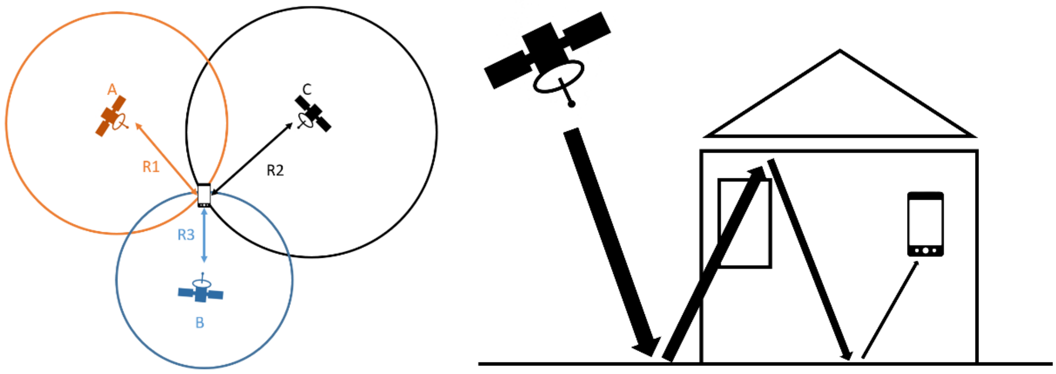


Fig. 0.1 GNSS problem

There are a multitude of technologies that try to solve this problem, but is not yet fully resolved. That is a motivation for writing this project. This document aims to establish good bases of knowledge for anyone who wants to know the location indoors. The project aims to give all the necessary aspects to understand such systems. Measurement algorithms, technologies and classification algorithms used in machine learning are described. All this provides the necessary to understand and analyze different classification algorithms knowledge. It also provides a context where to use them.

In this work, we focus on evaluating indoor positioning for smartphones. We have chosen to analyze standard WLAN networks to use as beacons. This type of system ignores the situation of beacons. The method of measurement is the Fingerprinting, which is based on mapping the area initially. Then, the position can be predicted from the data learned by classification algorithms. This solution prevents the installation of new devices because it reuses the previous infrastructure.

In this project, we hope to learn to use tools of machine learning. We believe it is very interesting and useful to know which are the main algorithms and how they work. From this knowledge, we will be able to understand the results and know when to use an algorithm or another. Since data management is something you learn with experience, we want to make the whole process and face all the challenges that entails. Consequently, we want to take our own experimental data and process them to make them useful.

This document does not want to leave out the commercial world. For that reason, we will try to make many references to real and commercial solutions. We want to make a useful, practical and accessible to all public work. Our intention is clear; we want to disseminate knowledge.

CHAPTER 1. BACKGROUND

1.1. Definition of indoor positioning

We can define an indoor positioning system as a technique that provides the continuous real-time location of objects or people within a closed space through measurements. There is still no standard that defines its operation and therefore there are many systems that provide different solutions to the same problem. The problem is that in a confined space GPS does not work.

The creation of GPS [1] enabled reliably and quickly locate devices that used its technology. GPS positioning techniques revolutionized and paved the way for the various GNSS [2] in the world today. Initially it gave military use, but later was opened to civil use. This allowed a lot of improving systems and services. Currently the phone industry has introduced this technology on their devices and have developed many applications that benefit from the use of positioning. Therefore, positioning is very important in smartphones.

Satellite guide is useful but requires direct view of the satellite. Although the satellite positioning system is being improved through the use of repeaters on the ground, still have problems indoors. Indoors, the satellite signal does not reach hard enough or is degraded by rebounds, a fact that causes failures in the device's location.

Techniques and systems have been developed to solve the problem by the need to service quality indoor positioning. In the following figure, we can see different technologies used for geolocation based on the uses.

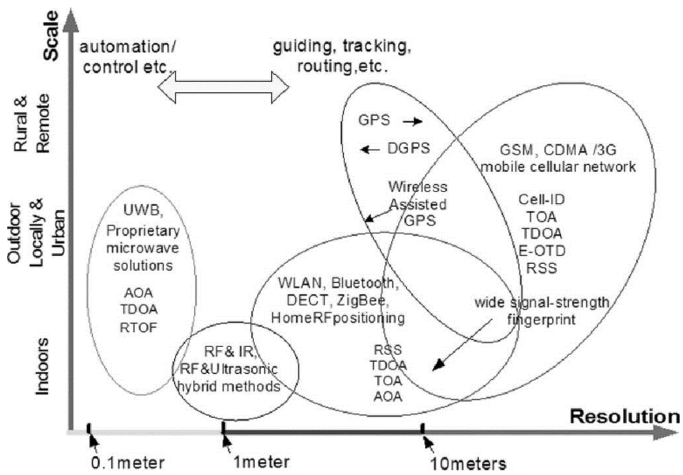


Fig. 1.1 Resolution of positioning technologies **[3]**

There are several indoor positioning technologies that provide high accuracy but require special systems and devices. The significant increase of mobile devices has contributed to an increase of Wi-Fi networks. Today we find Wi-Fi signals in most interiors of buildings in a city [4]. This fact makes the analysis of Wi-Fi signals for positioning may usually be used. [https://wigle.net/]

This work focuses on WLAN, because all smartphones have integrated this Technology and it is the most widely used technology to provide connectivity indoors. These advantages have made, indoor wireless positioning systems become very popular.

1.2. Measuring Principles

As has been mentioned, it is not easy to model the radio propagation in the indoor environment because of severe multipath, low probability for availability of line-of-sight (LOS) path, and specific site parameters such as floor layout, moving objects, and numerous reflecting surfaces. There is still no good model for indoor radio multipath characteristic and so the best option is to try to combine the various solutions.

Indoor location can differentiate different techniques that can be used with WLAN such as lateration, angulation, proximity detection and Location Fingerprinting. This project will focus on Location Fingerprinting (LF), but following other techniques are explained. If further information is required, please consult the following paper: [3].

1.2.1. Triangulation

Triangulation uses the geometric properties of triangles to estimate the target location. It has two derivations: lateration and angulation. The tracing and measurement of a series or network of triangles in order to determine the distances and relative positions of points spread over a territory by measuring the length of one side of each triangle and deducing its angles and the length of the other two sides by observation from this baseline.

- **Lateration** estimates the position of an object by measuring its distances from multiple reference points. So, it is also called range measurement techniques. Instead of measuring the distance directly using received signal strengths (RSS), time of arrival (TOA) or time difference of arrival (TDOA) is usually measured, and the distance is derived by computing the attenuation of the emitted signal strength or by multiplying the radio signal velocity and the travel time. Roundtrip time of flight (RTOF) or received signal phase method is also used for range estimation in some systems.

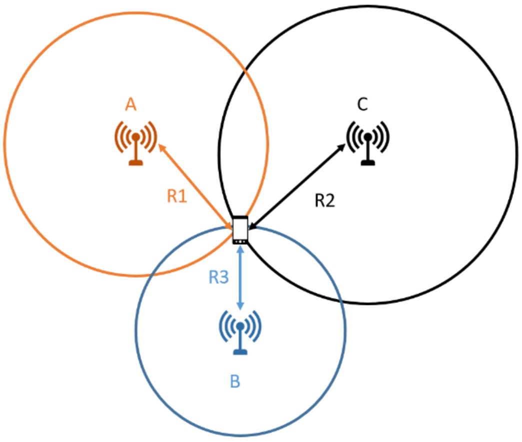


Fig. 1.2 Positioning based on TOA/RTOF measurements

- **Angulation** locates an object by computing angles relative to multiple reference points. In this survey, we focus on the aforementioned measurements in the shorter range, low-antenna, and indoor environment.

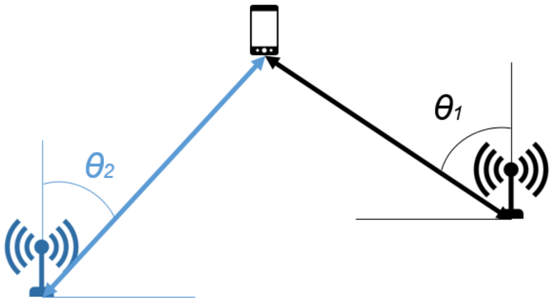


Fig. 1.3 Positioning based on AOA measurement.

1.2.2. Proximity

Proximity algorithms provide symbolic relative location information. Usually, it relies upon a dense grid of antennas, each having a well-known position. When a mobile target is detected by a single antenna, it is considered to be collocated with it. When more than one antenna detects the mobile target, it is collocated with the one that receives the strongest signal. This method is relatively simple to implement. It can be implemented over different types of physical media. In particular, the systems using infrared radiation (IR) and radio frequency identification (RFID) are often based on this method. Another example is the cell identification (Cell-ID) or cell of origin (COO) method. This method relies on the fact that mobile cellular networks can identify the approximate position of a mobile handset by knowing which cell site the device is using at a given time. The main benefit of Cell-ID is that it is already in use today and can be supported by all mobile handsets.

1.2.3. Location Fingerprinting

This document is focused in Location Fingerprinting. Location Fingerprinting is based on a database of pre-recorded measurements of network characteristics from different locations, denoted as fingerprints, a wireless client's location is estimated by inspecting currently measured network characteristics. Network characteristics are typically base station identifiers and the received signal strength.

1.3. Commercial technologies

This part of the project reviews the most commonly used positioning systems. Currently there are many indoor positioning technologies. Since the use of satellites, such as GNSS, to passive antennas, as in the case of RFID. As we shall see, none is perfect and fails to solve the problem of positioning interior completely.

1.3.1. GNSS

Global Navigation Satellite System (GNSS), are positioning systems using satellites emitting signals as reference points for positioning a receiver terminal. These systems work on frequency L. The L band, as defined by the IEEE, is the 1 to 2 GHz range of the radio spectrum. Using repeater beacons, it is possible to improve the received signal in GNSS and use the same system that GNSS positioning. This method, called Assisted GPS, can significantly improve indoor positioning. For users, it is one of the most natural solutions because this system is transparent to all applications that use satellite tracking. Furthermore, the service provider must supply all interior repeater beacons.

1.3.2. RFID

Radio-frequency identification uses electromagnetic fields to automatically identify and track tags attached to objects. It was initially designed as a smart tag, but currently can be used as a tracking system. The beacons are passive antennas that collect data terminal. By triangulation it is possible to locate the device, but typically the proximity to the positioning of the terminal is used. There are also models Fingerprinting. These beacons should not be more than 10 m from the terminal. In the event that the beacons are active sensitivity is higher. This system gives quite accurately, but unfortunately require specific terminals, so it is not appropriate for the case study of this project, which focuses on interior location using mobile phones. This model is often used in controlled locations such as hospitals, where it is necessary constantly monitor patients.

1.3.3. GSM/CDMA

GSM/CDMA has been used in cellular network communication. The frequencies used by this system has good propagation and building penetration. Generally, it falls in 850MHz, 900MHz, 1800MHz, and 1900MHz bands. Using GSM / CDMA for localization is not new. Normally triangulation is used for this purpose, although not giving much precision. The problem arises when trying to position indoors. As mentioned, electromagnetic signals undergo attenuations and rebounds that hinder triangulation. Since the triangulation does not work well indoors, other methods are being tested. Fingerprinting studies prove that give better results [5]. Unfortunately, GSM/CDMA is heavily patented, so it is hard to do modification or extensions based on GSM/CDMA which limits the future development on it.

1.3.4. UWB

Ultrawideband (UWB) is a technology in the range of the PAN (personal area network). It allows very large information packets (480 Mbits / s) achieved over short distances of a few meters. Current wireless USB devices are implemented with UWB. UWB uses a higher bandwidth of 500 MHz or 25% of the center frequency, according to the FCC (Federal Communications Commission). The power consumption is low. Another advantage of UWB is its immune to multi-path problems. The positioning error of this technology is very low, but unfortunately today's mobile phones don't have incorporated such technology. But there are signs that this technology will soon be incorporated into mobile devices [6].

1.3.5. WLAN (IEEE 802.11)

Wi-Fi is one of the most used wireless technology. Therefore, there are plenty of base stations that can be used as a reference point. It uses two licensed-exempt bands: 2.4 GHz, and 5 GHz. Both mobile phones and laptops, have a connection to this network. There are several location systems that use WLAN. That means the infrastructure cost and user device cost can be very low. Additionally, Wi-Fi based localization can be easily adopted by buildings and users. Some use special beacon for this purpose, and improve outcomes of location. The disadvantage of this technology is that it is not generic and therefore can only be used wherever is installed. In our case, we use the other case. Positioning by generic devices. Such systems use signal strength, no special time markers.

1.3.6. Bluetooth (IEEE 802.15)

Bluetooth is a personal area network standard. It was designed to connect mobile devices wirelessly. It uses the same bands as Wi-Fi does (2.4 GHz and 5 GHz). It is widely used for short distance communication like headset, mobile phones

The Bluetooth transmission power is very low and therefore the coverage of Bluetooth is shorter than Wi-Fi and other WLAN technology. Consequently, Bluetooth is not appropriate to cover a large area.

1.3.7. Other

Other technologies are rarely used, but in this section, we want to mention.

1.3.7.1. Visible Light Communications

Information can be transmitted through light. There is no visible light and visible light. Currently there are companies using lamps malls to transmit signals to mobile phone cameras. The phone would be able to capture those signals and position in the room. [7] [8]

1.3.7.2. Artificial Vision

There are companies that are starting to use digital cameras and image processing for people recognition, tracking and positioning. In that way, it would be possible to position an individual. [9]

1.3.7.3. NFER

Near-field electromagnetic ranging is an emerging RTLS technology that employs transmitter tags and one or more receiving units. Operating within a half-wavelength of a receiver, transmitter tags must use relatively low frequencies (less than 30 MHz) to achieve significant ranging. Depending on the choice of frequency, NFER has the potential for range resolution of 30 cm and ranges up to 300 m. This kind of signals allow positioning in harsh environments such as factories with many metallic elements or nuclear power plants. [10]

1.3.7.4. Mobile sensors

The sensors of a mobile phone can help positioning. The electromagnetic sensor can give guidance, the gyroscope can tell us if they are climbing stairs, even if you're in an elevator. The sensors are often used as support for other technologies.

1.3.8. Solutions

It is very interesting to consider the variety of commercial solutions exist on the market. For this reason, it has been included a list of companies working in indoor positioning Annex. In total, 47 companies have been evaluated. The following chart shows the percentage of firms using each of the technologies.

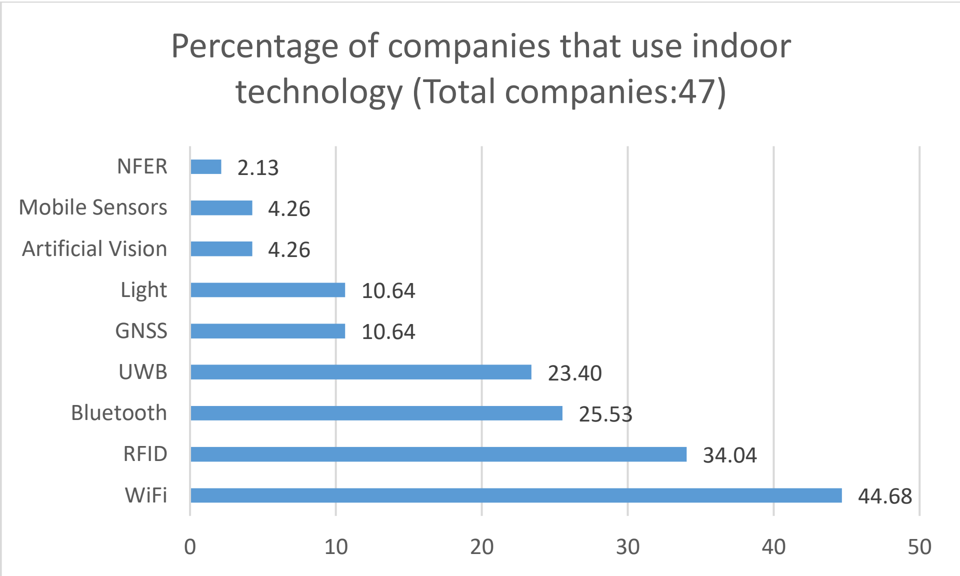


Fig. 1.4 Percentage of companies using indoor technologies

As shown, WLAN technology is the most commonly used, followed by RFID, Bluetooth, UWB and GNSS. Other technologies are rarely used. If we analyze the advantage of Wi-Fi is its widespread use in cities. As we can see in the table, it is a case in which reuses existing infrastructure and it is not necessary to build a dedicated.

Table 1.1 Wireless technologies used for indoor localization **[11]**

Wireless Technology	Range	Dedicated Infrastructure	Power Consumption	Disadvantages
GPS repeater	30 m	Yes	low	high variance signal
Wi-Fi	35 m (indoor)	No (for most places)	high	high variance signal
Bluetooth	10 m	Yes	low	Cover range is limited
UWB	few meters	Yes	low	Cover range is limited
RFID	1~10m	Yes	low	Cover range is limited

1.4. WLAN Mapping

For some time, there are people dedicated to creating open WLAN maps in cities around the world. By their own devices, whether a laptop, a tablet or a smartphone, there are applications that allow capture WLAN and locate them in specific coordinates. Systematic capture WLAN along a territory can generate maps. This practice is called Wardriving. There are different types of websites where upload these data: websites like WiGLE [4], openBmap [12] or Geomena [13]. These maps can be useful because known network IDs can be used as a then a geolocation system, an alternative to GPS when a device that has no vision positioning satellites. Hotspots become benchmarks and using different techniques we know our location.

Some examples include Place Lab [14], Skyhook [15], Navizon [16], openBmap and Geomena. Navizon and openBmap combines information from Wi-Fi and cell phone tower maps contributed by users from Wi-Fi-equipped cell phones. In addition to location finding, this provides navigation information, and allows for the tracking of the position of friends, and geotagging. Google has also made use of this system and has decided to systematically collect these data with their Street View cars. In some countries, it has created controversy because privacy laws are violated. [17]

Initially, it was considered studying interior positioning Google given its widespread use. But their deployment was not yet in this country and forces you to use their API (application programming interface) without showing the algorithms used. That's why this project was aimed at an analysis of the different algorithms.

1.5. Classification Location Fingerprinting

The technology can be classified according to different taxons per different criteria. The proposed taxonomy is built around eleven taxons listed with definitions bellow. The taxons scale, output, measurements, and roles describe general properties of LF systems. The map of the complete taxonomy is in Annex. If more doubts about the taxonomy, see bibliographic source [18].

Table 1.2 Description of the main taxons

Taxons	*Definition*
Scale	Size of deployment area.
Output	Type of provided location information.
Measurements	Types of measured network characteristics.
Roles	Division of responsibilities between wireless clients, base stations, and servers.
Estimation Method	Method for predicting locations from a radio map and currently measured network characteristics.
Radio Map	Model of network characteristics in a deployment area.
Spatial Variations	Observed differences in network characteristics at different locations because of signal propagation characteristics.
Temporal Variations	Observed differences in network characteristics over time at a single location because of continuing changing signal propagation.

Sensor Variations	Observed differences in network characteristics between different types of wireless clients.
Collector	Who or what collect fingerprints.
Collection Method	Procedure used when collecting fingerprints

The project was carried out in a University campus and tries to position the device inside the building, indicating the point, the room or the floor. It is a discreet and not numerically. The system is based on a Radio Map which has been picked up by an Administrator which collects the signal power variations in each of the points. The system is trained using different algorithms and can classify a sample in a given location. Any terminal would be able to read samples and the server can tell the position of the map.

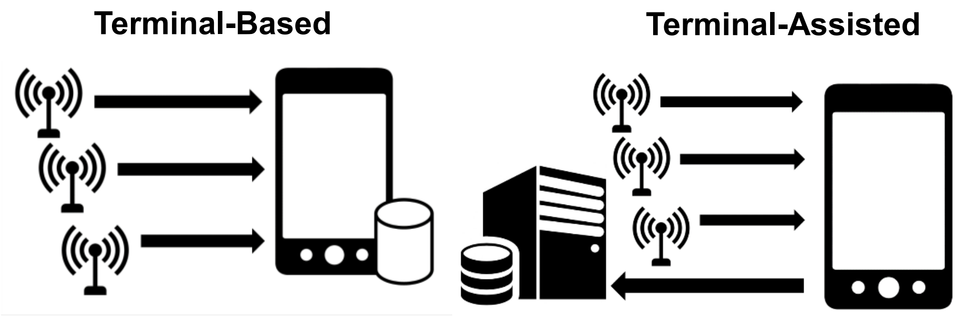


Fig. 1.5 Terminal-Based vs Terminal-Assisted

The following table tries to summarize the taxonomic characteristics of this project:

Table 1.3 Taxonomic characteristics of this project

Taxon	*Values of the project*
Scale	Building, Campus
Output	Descriptive (plants, room, points)
Measurement	Terminal-assisted/terminal based
Estimation Method	Determinist/probabilistic
Radio Map	Empirical
Spatial Variations	Fingerprint filtering
Temporal Variations	History of Measurements
Sensor Variations	Mapping
Collector	Administrator
Collection Method	Known location Point spatial property Fixed number of measurements

1.6. Conclusions of the chapter

This chapter defined the principles used in positioning. Once these principles are known, we can understand that current systems do not work properly indoors. The mention of the technologies used currently provides a global vision. The increased use of WLAN for indoor positioning is appreciated. Also, several initiatives have been mentioned fingerprinting using WLAN for positioning. Precisely this project explores different classification methods using this technology. This chapter provides the necessary knowledge to understand the context of indoor positioning.

CHAPTER 2. MACHINE LEARNIG ALGORITHMS

Experience and intelligence are needed to make predictions. Machine learning is a field of computer science that studies how to teach a machine to learn. These algorithms try to anticipate a response from prior learning. Depending on the system of learning algorithms can be divided into three types: supervised learning, unsupervised learning, reinforcement learning.

2.1. Supervised learning

In this case the algorithms are based on a set of data previously entered. Usually we predict variable (dependent variable) which is obtained from the learned variables (independent variables). It is called supervised because learned variables are from a set of data selected and classified by a human agent. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

2.2. Unsupervised learning

This algorithm does not have any target or outcome variable to estimate. Usually, it is used for clustering. Data is classified in different clusters or cells. Grouping information is widely used in business issues. It helps to better understand the data. Examples of Unsupervised Learning: Apriori algorithm, K-means.

2.3. Reinforcement learning

This type of algorithm forces the machine to take specific decisions. The machine is exposed to the environment and trains on their own using the system of trial and error. The machine learns from the experience and try to improve the prediction. Example of Reinforcement Learning: Markov Decision Process

2.4. Algorithms analyzed

Then we will see a list of the most common machine learning algorithms.

2.4.1. Naive Bayes

Based on Bayes' theorem, this is a classification technique with an assumption of independence between predictors. It assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c)\times P(x_2|c) \times \cdots \times P(x_n|c)\times P(c)$$

- P(c|x) is the posterior probability of class (target) given predictor (attribute).
- P(c) is the prior probability of class
- P(x|c) is the likelihood which is the probability of predictor given class
- P(x) is the prior probability of predictor

Below we will try to understand this algorithm using an example.
Imagine you have a table with the higher power device detected based on the second floor of a building.

Table 2.1 Data set example

max power device detected	2nd floor
MAC1	No
MAC2	Yes
MAC3	Yes
MAC1	Yes
MAC1	Yes
MAC2	Yes
MAC3	No
MAC3	No
MAC1	Yes
MAC3	Yes
MAC1	No
MAC2	Yes
MAC2	Yes
MAC3	No

Now, classify whether it is on second floor or not based on the detected devices.
First, convert the data set to frequency table:

Table 2.2 Frequency table

Frequency Table		
max power device detected	No	Yes
MAC1	2	3
MAC2	0	4
MAC3	3	2
All	**5**	**9**

Second, create Likelihood table by finding the probabilities like MAC2 probability = 0.29 and the probability of being on the second floor is 0.64.

Table 2.3 Likelihood Table

Likelihood Table				
max power device detected	No	Yes		
MAC1	2	3	= 5/14	0.36
MAC2	0	4	= 4/14	0.29
MAC3	3	2	= 5/14	0.36
All	**5**	**9**		
	= 5/14	= 9/14		
	0.36	0.64		

At last, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

using above discussed method, we can predict whether we are on the 2nd floor if device detected is MAC1:

$$P(Yes \mid MAC1) = P(MAC1 \mid Yes) * P(Yes) / P(MAC1)$$

Here we have,

$$P(MAC1 \mid Yes) = 3/9 = 0.33$$
$$P(MAC1) = 5/14 = 0.36$$
$$P(Yes) = 9/14 = 0.64$$

Finally,

$$P(Yes \mid MAC1) = 0.33 * 0.64 / 0.36 = 0.60,$$

which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes. In the case of Bayes Network, events can be interrelated. In our case, we will use:

weka.classifiers.bayes.NaiveBayes [19]
weka.classifiers.bayes.BayesNet

2.4.2. Rule Lerner

The method is based on the recursive solving a problem by dividing it into two or more subproblems of the same type or similar. The process continues until they become simple enough to be resolved directly. In the end, the solutions to each of the subproblems are combined to give a solution to the original problem.

Such algorithms are also known as "Divide and conquer" Algorithms, as in popular culture reference to a saying that involves solving a difficult problem, dividing it into simpler parts as many times as necessary, is done until resolution becomes obvious. The solution of the main problem is built with the solutions found. Some examples of such algorithms are ZeroR, OneR [20] or JRIP. In our case, we will use the following algorithm:

weka.classifiers.rules.JRip [21]

2.4.3. Decision Tree

Decision Tree is a supervised learning algorithm that is mostly used for classification problems. The algorithm splits the population into two or more homogeneous sets. This is done based on most significant independent variables to make as distinct groups as possible. It works for categorical and continuous dependent variables.

To carry out our tests we decided to use the J48:

weka.classifiers.trees.J48 [22]

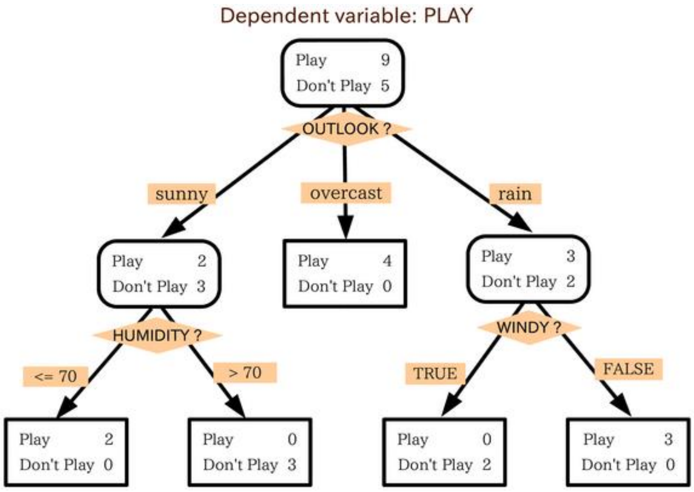


Fig. 2.1 Decision Tree example **[23]**

2.4.4. KNN (K- Nearest Neighbors)

It is more widely used in classification problems. K nearest neighbors stores all available cases and classifies new cases to the class is most common amongst its K nearest neighbors. K nearest neighbors are measured by a distance function. These distance functions can be Euclidean, Manhattan, Minkowski and Hamming distance. Hamming is used for categorical variables. K is the number of neighbours. If K = 1, then the case is assigned to the class of its nearest neighbor.

In our case, we decided to use the following algorithm of the collection:

weka.classifiers.lazy.IBk [24]

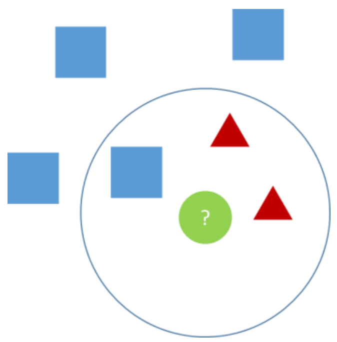


Fig. 2.2 KNN example

2.4.5. SVM (Support Vector Machine)

SVM is a classification tool. Each feature is represented as a dimensional space and every data item as point in these hyper-planes with the value of each feature being the value of a coordinate. Separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class. In general, the larger the margin the lower the generalization error of the classifier, since the larger the margin the lower the generalization error of the classifier. In Weka, we have several algorithms that allow us to test, such as:

weka.classifiers.functions.SMO. [25] [26]

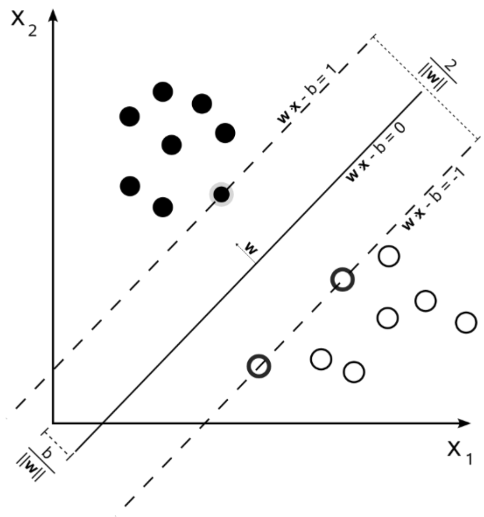


Fig. 2.3 Support Vector Machine **[27]**

2.4.6. Neural network models

The concept of artificial neural networks is inspired in the examinations of humans' central nervous systems. Neural models are based on the principle of non-linear, distributed, parallel and local processing and adaptation. The model is built from simple nodes, known as “neurons” or "processing elements". This units are connected to form a network which mimics a biological neural network. The following shows an example of a neural network

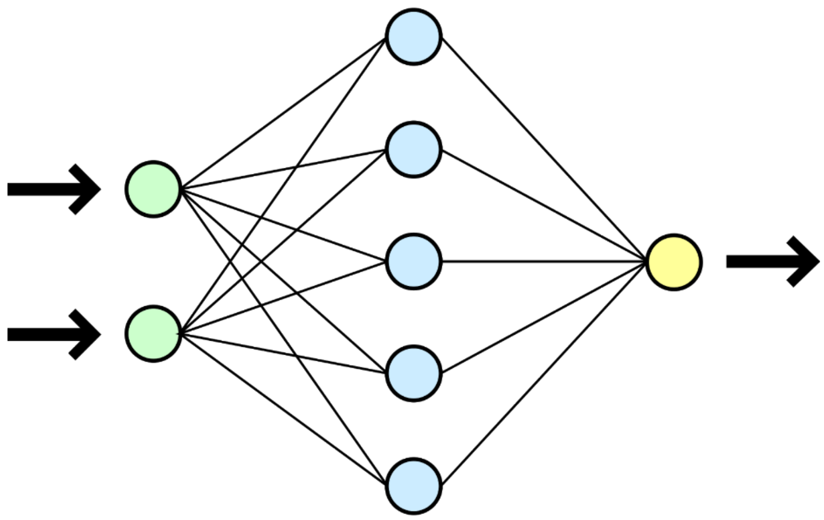


Fig. 2.4 Neural model graph **[28]**

Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another. A neural network contains sets of adaptive weights, parameters that are tuned by a learning algorithm. The adaptive weights can be thought of as connection strengths between neurons, which are activated during training and prediction. It has the capability of approximating nonlinear functions of their inputs. Weka has several algorithms. The next, it would be one of them:

weka.classifiers.functions.MultilayerPerceptron [29]

2.5. Conclusions of the chapter

Machine learning is a branch of artificial intelligence that is studied in computer science which aims to develop techniques that allow computers to learn. In our case, we are concerned that the machine is able to learn the difference between a certain position of map and another, based on the received power of multiple transmitters in the area. So, what is desired is that the machine can classify.

This chapter described classification algorithms that have been discussed below. It is important to consider the functioning of these algorithms to understand the results of tests performed in the following section. Machine Learning is an experimental science and the selection of an algorithm depends on the characteristics of the data and use that are given. In the case of changing any of these properties should be retested.

For this project, we will use Machine Learning algorithms to locate the position of a user. The algorithm compares the power values Wi-Fi at a point that has been previously analyzed. Our project is based on Supervised learning as it uses a database that has been previously trained. Our goal is to clarify which of the algorithms we have explained is the best to perform this task.

The algorithms evaluated in this project are:

- weka.classifiers.bayes.NaiveBaye (Naive Bayes)
- weka.classifiers.bayes.BayesNet (Naive Bayes)
- weka.classifiers.rules.JRip (Rule Lerners)
- weka.classifiers.rules.ZeroR (Rule Lerner)
- weka.classifiers.trees.J48 (Decision Tree)
- weka.classifiers.lazy.IBk (K- Nearest Neighbors)
- weka.classifiers.functions.SMO (Support Vector Machine)
- weka.classifiers.functions.MultilayerPerceptron (Neural network models)

CHAPTER 3. Experimental Framework

3.1. Scenario

Then we proceed to describe the project scenario. Samples of this project have been obtained in the building of the Polytechnic University of Castelldefels. The tests were conducted from 0 floor to the 3rd floor. In the next picture, the area analyzed is shown.

In order to take the data a mobile phone Samsung Galaxy SII, which employs an Android operating system, was used. The application Wi-Fi Maps Light captured the MAC [30] of wireless devices that were radiating around us. The transmission power was also captured, so we could create a map of the place.

Here are all the points where the data was collected. In each of these points were analyzed at least 10 samples. Samples were taken looking in different directions, so the effect of rebounds was taken into account.

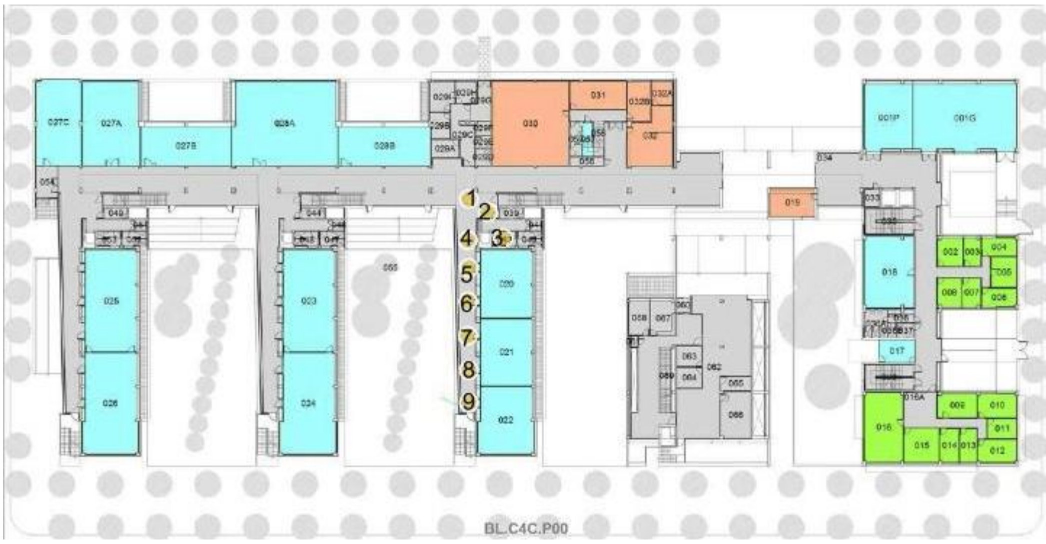


Fig. 3.1 Scenario map

The sampling points are represented with a yellow circle. There is a total of 9 points in each of the plants. The points have been identified with a number. For testing it has been encoded point name as follows:

pl0_p01

where in PL0, 0 corresponds to the floor (0,1,2,3) and p01, 01 corresponds to point number of the previous figure (01-09). (To view the plans of other plants look annex)

The following bar chart shows the samples collected per plant:

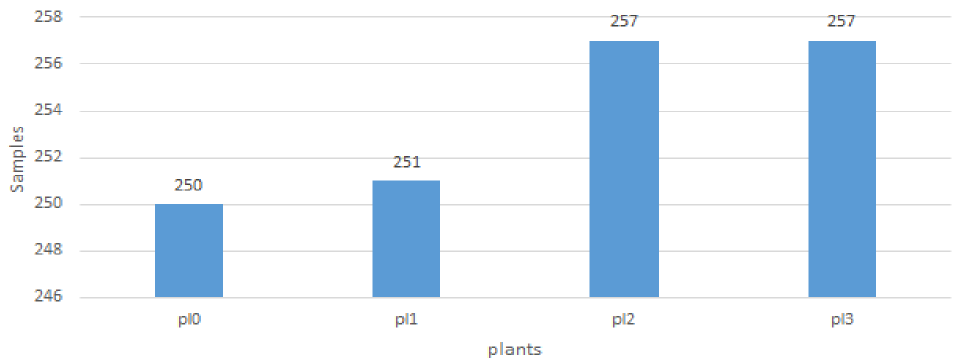


Fig. 3.2 Number of samples per plant

We can see that we have collected about 250 samples per plant. Because we have a lot of samples the following chart helps us see some more detail with samples collected by point:

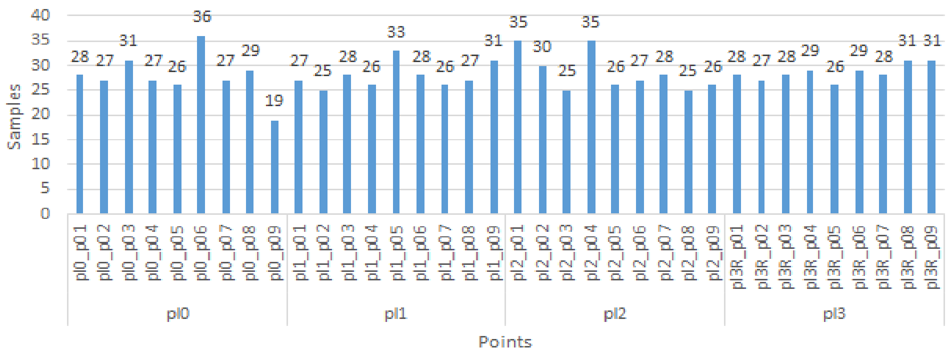


Fig. 3.3 Number of samples per point

The number of samples taken at each point varies between 29 and 36. This is because the samples were taken manually and no automated control of data collection.

Here we will focus on the content of the samples. In the collected sample, it was considered the BSSID (Basic Service Set Identifier) [31] received by each of the devices. We start from the premise that the power difference of each of the signals can help geolocation. The following graph shows the average intensity of each device (named for its MAC [32]) on each floor.

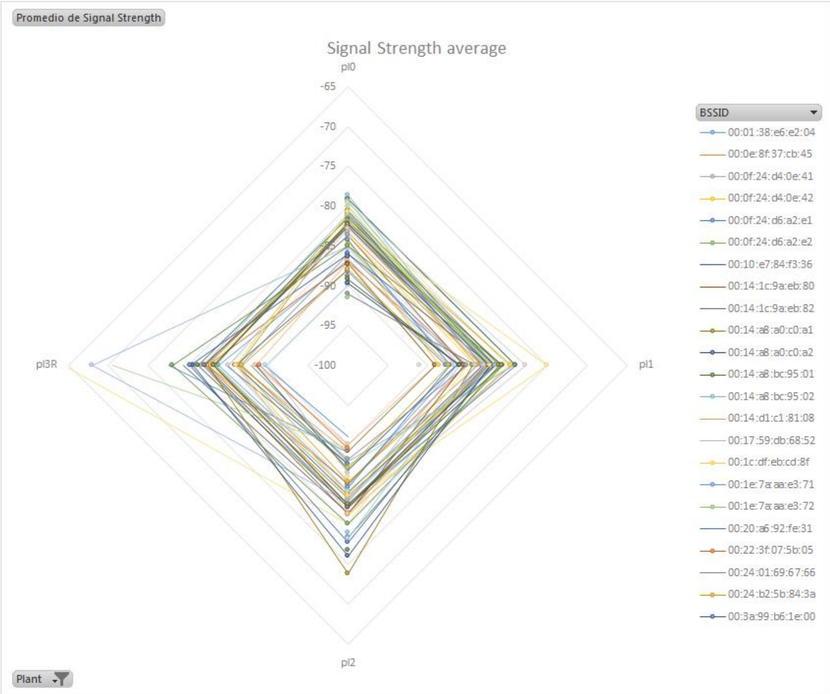


Fig. 3.4 MAC average intensity per plant

The figure above shows the average received power of each device on each floor of the building. There are clear differences on each floor. These differences can help us locate the subject on one level or another in a decision tree, for example. We also see that there are signals that keep about its value in all plants and do not give us differentiating information. These signals add unnecessary calculations that slow down the system and can mislead us. They can be considered noise and so we have to look deeper into the problem to filter properly.

3.2. Data Tools

To carry out this section we have used python programming for data manipulation and Weka platform [33] (a collection of machine learning algorithms for data mining tasks). We have analyzed and compared classification algorithms to determine the position of mobile phones using the power of wireless signals surrounding it.

Weka has a visual interface developed in JAVA. It that facilitates analysis and comparative data mining algorithms. Weka Explorer is used to try different datasets and algorithms with a clear and simple interface. This helps us visualize the results and understand what each algorithm do. Weka Experimenter is used to execute various algorithms loop on the same dataset and compares between them. The other sections (KnowledgeFlow, simple CLI) are not used in this project. To learn more about Weka, you can watch this course. [34]

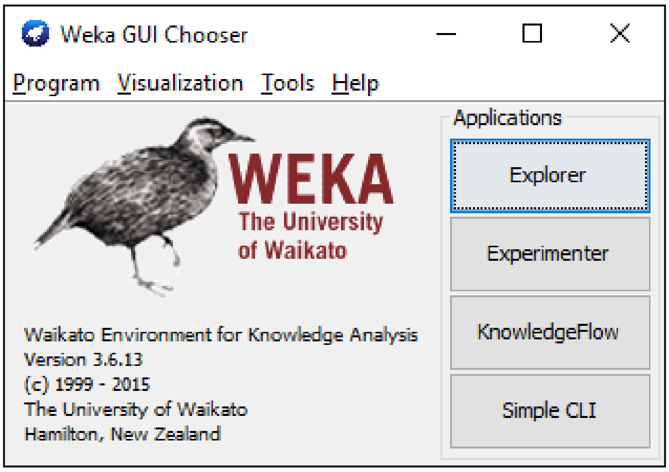


Fig. 3.5 Weka GUI Chooser

3.3. Experimentation

To experiment with algorithms must select a database of training and another testing dataset. To get a first idea of the percentages of success of the algorithms, it has made the following graph.

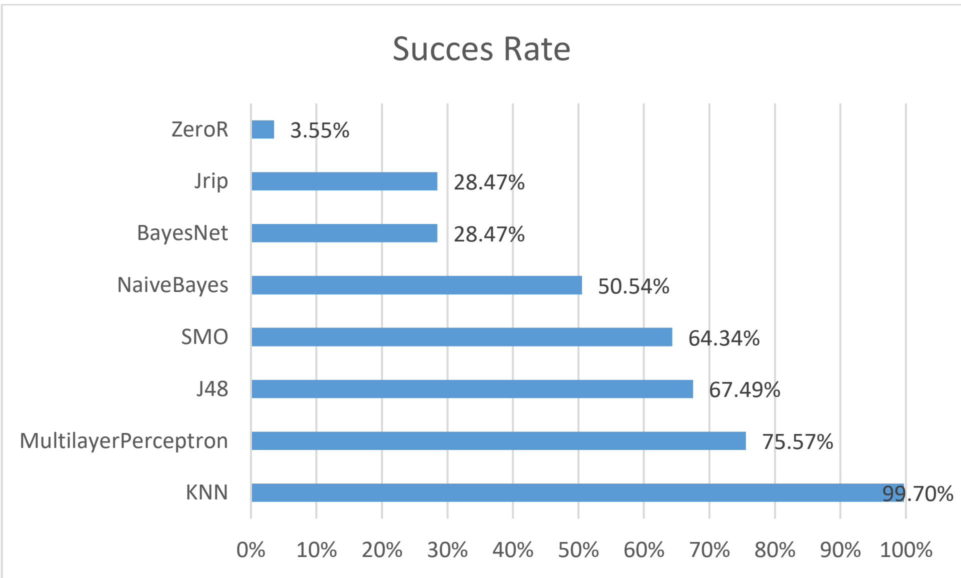


Fig. 3.6 Success Rate

The figure above shows the percentages of success that has earned each classifier using the same dataset for training and testing. In this case, if we use the same dataset, the results will be much better than the results obtained in a real environment. Under these testing conditions, KNN gets the best results, followed by a neural method. Since the KNN has obtained such positive return results, we hypothesize that this algorithm is a good candidate for use in indoor positioning.

3.3.1. Algorithm Properties

Each algorithm has different variables that can be modified to tune their results. Since Knn has shown the best results in the previous test, we will experience several of its properties.

KNN classification based on samples learned. This means it must have a data set stored in memory and whenever it decides to classify a sample, see its entire database.

Weka provides the functionality that allows to modify several parameters of the algorithm. It has been chosen to modify the number of neighbors. Neighbors are taken by the algorithm to determine the class of the sample. The more neighbors in the same class there, it is more likely to be of that class. It can also vary the weight of the neighbors per the distance.

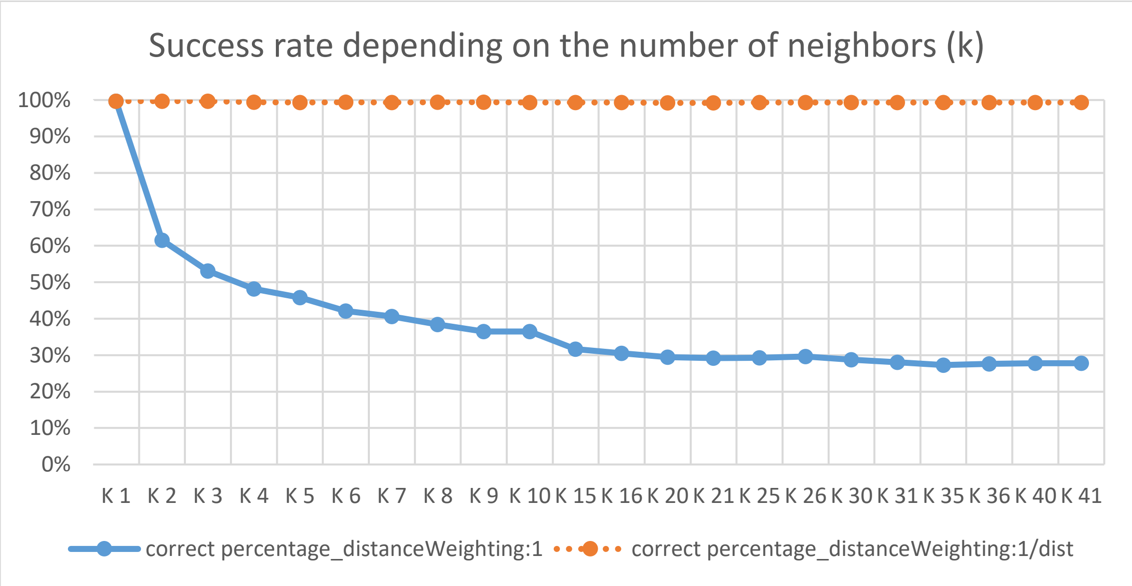


Fig. 3.7 Success rate depending on the number of neighbors (k)

The horizontal axis of the graph shows the variation (k), where k is the number of neighbors taken. Ideally this value should be odd to avoid problems in decision making. For example, the case in which there are the same number of neighbors classified as a class and another. In this experiment, we have also taken even values of K. The vertical axis shows the percentage of success of each run.

In Figure, we find a set of blue values, representing the results of increasing k regardless of the distance of each neighbor with the sample to be classified. We see that the percentage of success is becoming less because values of several classes are taken and that makes classification fail. On the other hand, if we take a weight inversely proportional to the distance, we weigh less samples are away and will weigh more nearby. Thus, the success rate remains high.

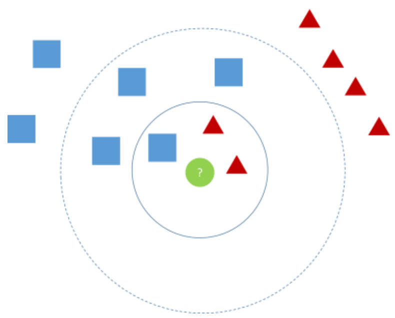


Fig. 3.8 KNN classification interference

As one can see in the example of the previous figure if too wide the number of neighbors, more elements of other classes that may affect the classification are added.

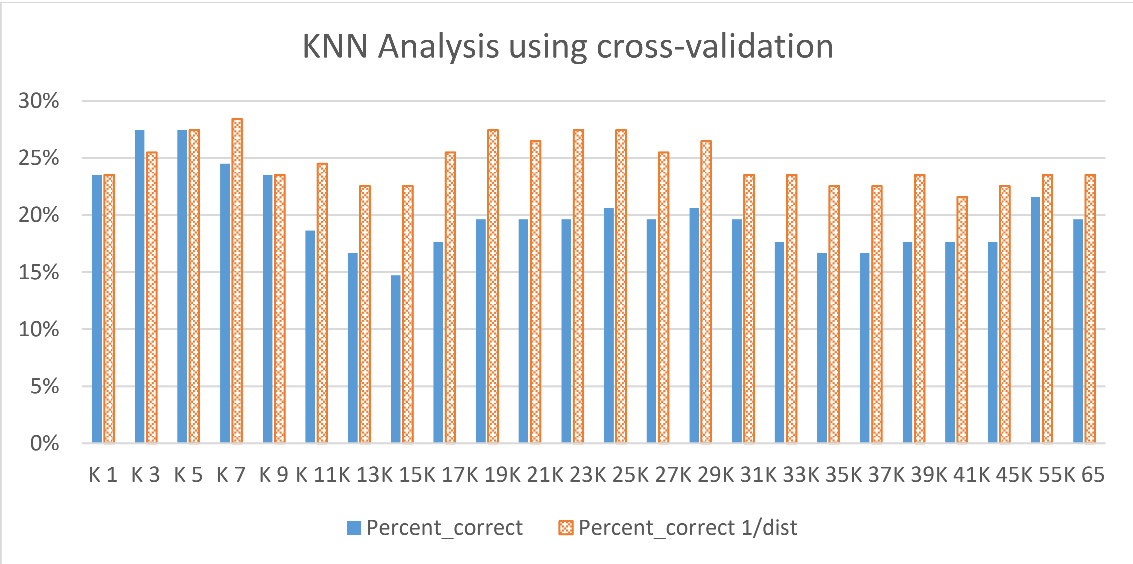


Fig. 3.9 KNN Analysis using cross-validation

According to stated earlier, if we classify with the same training dataset we obtain idealistic values. To avoid that and get real values a set of cross-validation tests, which provides more realistic values were executed. This graph is still seen when one takes into account the distance of samples continues to give better results, but the value of K does not have to improve outcome. We will see later that KNN is heavily affected by noise. The above results are proof.

3.3.2. Attribute selection

In the previous section (3.2) we said that too many signals can cause noise and so we use methods of selection of optimal attributes. Weka has several algorithms that eliminate less relevant attributes. We used these algorithms to reduce noise and processing load.

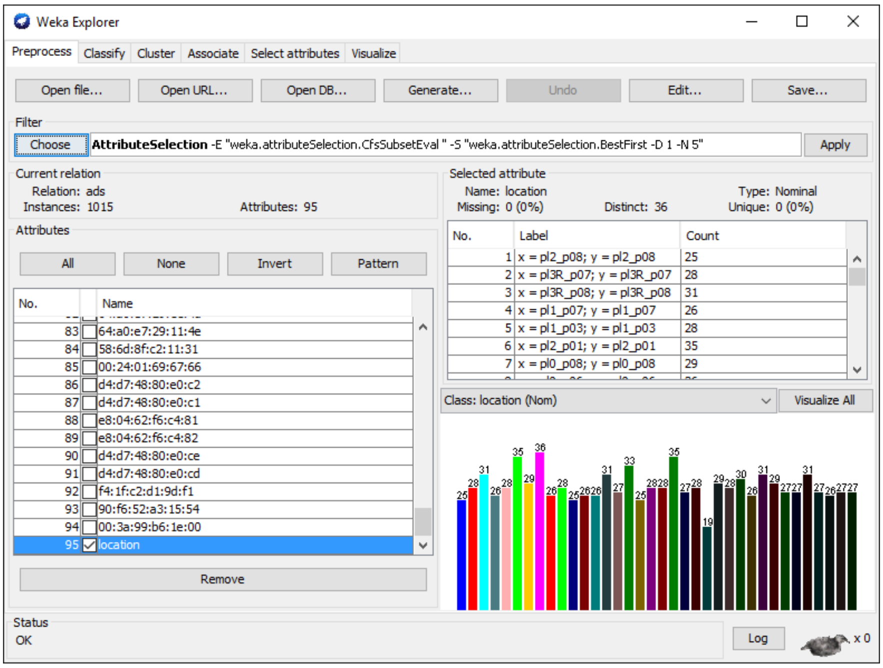


Fig. 3.10 Weka explorer

The figure above shows the preprocessing tab of this program. This panel lists the data attributes. Choosing one, a graph with the distribution of data is generated according to the selected attribute. To select the filter and pre-processing attributes, we must look at the filter button to select the type of algorithm to be applied to the samples. The algorithms selected are as follows.

Table 3.1 KNN classification results

Attribute Selection Method	number of attributes	Correctly Classified Instances	success rate
Cross-validation 10 folds	95	259	25.5 %
Cross-validation 10 folds weka.filters.supervised.attribute.AttributeSelection -E "weka.attributeSelection.CfsSubsetEval " -S "weka.attributeSelection.BestFirst -D 1 -N 5"	48	247	24.3 %
Cross-validation 10 folds weka.filters.supervised.attribute.AttributeSelection -E "weka.attributeSelection.CfsSubsetEval " -S "weka.attributeSelection.BestFirst -D 1 -N 5"	10	205	20.2 %

We used two methods of selecting attributes. In either cases the success rate exceeds the initial classification. We can rule out that the low success rate is due to the noise caused by the large number of attributes. In return they have considerably reduced the number of attributes to be analyzed. CfsSubsetEval algorithm has reduced the success rate to 5%, something not acceptable. But in the case of ClassifierSubsetEval it has managed to halve the number of attributes and has only reduced by 1% success rate.

3.3.3. Preprocessing Dataset

The attribute reduction results have not improved significantly. We infer that the poor performance of classification is not due to an excessive number of attributes. Therefore, we will consider other options. We believe that if the spacing between points is increased it may give you a better success rate. If we have two points close together it is very likely that the measured signals are very similar and the system is not able to differentiate one point of another. On the other hand, if the points are far apart, the measured signals will be different enough to easily distinguish the two points. To increase this distance, we have decided to remove parts of the system. If we remove the even points, we move away the distance between adjacent points.

With the intention of improving the success rate, we reduced the requirement of the system by reducing points. We eliminated some the points per plant and observed the results. For this task, we used python to eliminate the points that interest us. We have used the Anaconda package [35]. The aim was to create a .arff file per plant. (Another option is to use the Weka platform - where you can select the points we are interested). In the annex, you can see the python code.

We can sort the data by plants and points of each plant. From the Weka Explorer, we select the tab Classify. For this test, we selected KNN classifier. Previously we used the "Use training set" option and gave us a result close to 100% accuracy. This is because it uses the same file system to train and classify. Now we will use Cross-Validation dividing the values in folds and randomizing values for testing and training. The division into folders is not easy. In the case of Stratified Cross-Validation, the division tries to maintain the same distribution of classes in all folders. This method provides more realistic results.

The following chart displays the classification of Weka. As shown, you can choose several options to perform the tests. This is shown as choosing the cross-validation 10 Folds. Also, you can see the classifier.

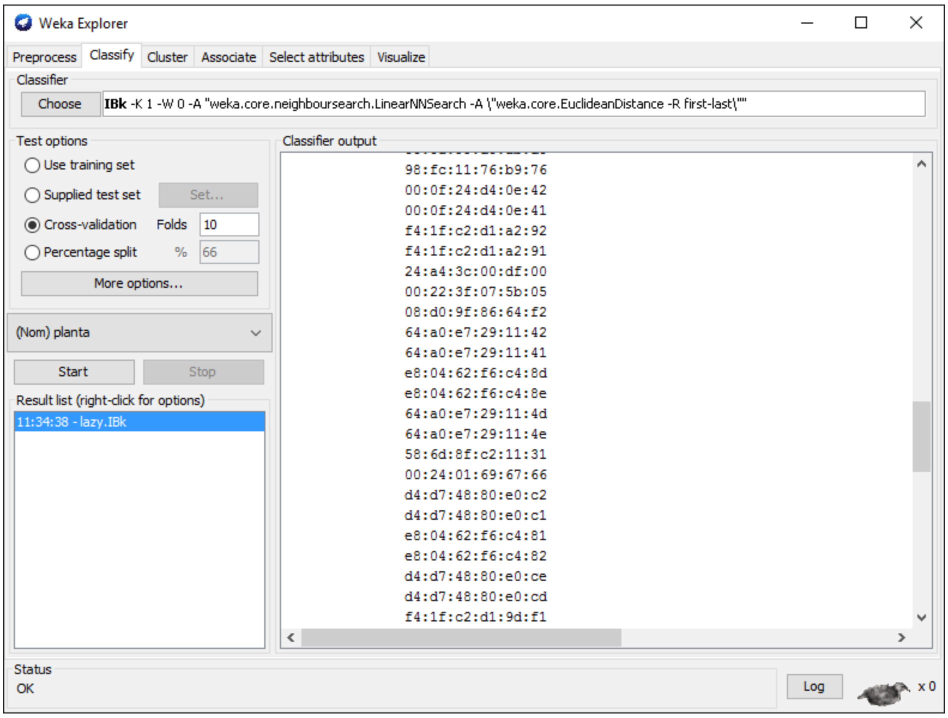


Fig. 3.11 Weka KNN classifier using Cross-Validation

The figure below shows the results after the execution of weka.classifiers.lazy.IBk method. The section called "Run information" gives all the information concerning the execution of the algorithm. We can see the name of the algorithm executed along with the input parameters. We can also see the database and filters. In addition, it indicates the number of instances and attributes of the database. Part of summary shows the main results. "Detailed Accuracy By Class" separates the accuracy results in different classes. "Confusion Matrix" is a great tool to know the number of successes and failures of a classifier.

```
=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation: all_plantasR+L
Instances: 1015
Attributes: 96
[...AtributesName...]
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds
```

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        854               84.1379 %
Incorrectly Classified Instances      161               15.8621 %
Kappa statistic                         0.7885
Mean absolute error                     0.0806
Root mean squared error                 0.281
Relative absolute error                21.4916 %
Root relative squared error            64.8993 %
Total Number of Instances            1015

=== Detailed Accuracy By Class ===

        TP Rate   FP Rate Precision Recall F-Measure ROCArea  Class
          0.879     0.046     0.866    0.879    0.873     0.915       pl2
          0.883     0.034     0.897    0.883    0.89      0.925       pl3
          0.765     0.071     0.78     0.765    0.773     0.844       pl1
          0.836     0.06      0.82     0.836    0.828     0.896       pl0
W.Avg.   0.841     0.053     0.841    0.841    0.841     0.895
(Weighted Average)

=== Confusion Matrix ===

   a    b    c    d   <-- classified as
 226   13   13    5 |   a = pl2
  11  227    8   11 |   b = pl3
  22    7  192   30 |   c = pl1
   2    6   33  209 |   d = pl0
```

Fig. 3.12 KNN results using Cross-Validation (using Weka)

The classification for each floor is significantly better than the point classification we made earlier:

Correctly Classified Instances (point classification)	259	25.5172 %
Correctly Classified Instances (plant classification)	854	84.1379 %

This is because the separation is much higher. This is confirmed when separate points of each plant. After these results, we decided to evaluate the resolution of our system. We carry out tests in which we eliminated points to increase the distance between adjacent points. In the following chart the results of these experiments are shown. The vertical axis represents the average success rate of all tested algorithms. On the horizontal axis, the experiments performed and the number of points in parentheses are exposed. The color of each line represents the floor.

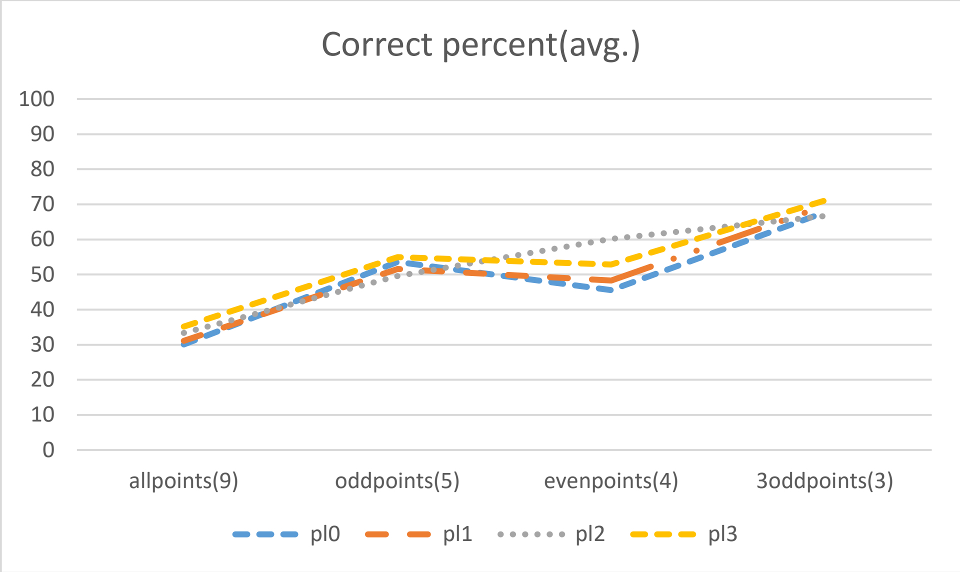


Fig. 3.13 Success rate according separation of classes

Then we describe the variables X axis

- Allpoints (9): We selected all points on all floors to the test. A total of 9 points per plant.
- Oddpoints (5): We have chosen a total of 5 points odd numbered.
- Evenpoints (4): We have chosen a total of 4 points even numbered.
- 3oodPoints (3): We chose the 3 most physically separated odd points.

In this graph, we can see how improved the success rate on each floor as separate points. To obtain these data, the samples were separated by plant and evaluated with all algorithms of this project. The average result of success is what is shown. Thus, we can see that the average value of success increases regardless of the type of algorithm. It must be said that not only the number of points affect the classification, also the arrangement between them.

We can also see that the values do not reach the rates that we have shown above. This is because the graph shows an average success rate of all the results of all algorithms.

3.3.4. Main concepts for analysis

In the results of the previous section, we saw many values and concepts that had not been defined. Before going further, we discuss the meaning of several important parameters.

3.3.4.1. Confusion Matrix

The confusion matrix is one of the clearest ways to understand the results of a classifier. Each column represents a value classification. Each row equals the actual value of the classified sample. The first value of the array is 226. This means that 226 samples were correctly classified because they represent the values classified as second floor and its real value is second floor. The same happens in the second column and the second row. The diagonal of a confusion matrix contains correctly classified samples. If we continue the analysis, the first row shows that the remaining original values of the 2nd floor have been classified as follows: 13 samples on the third floor, 13 on the first floor and 5 on the ground floor.

3.3.4.2. TP Rate & FP Rate

It has been the distribution of the successes and failures of classification in a confusion matrix. True Positive Rate (TO Rate) is the relationship between successful values and total values of that class. Matches the value of Recall. Below it is shown how to calculate the TP Rate. It is calculated for the second floor of the previous result.

$$\text{TP Rate} = \Sigma \text{ True positive}/\Sigma \text{ Condition positive}$$
$$\text{TP Rate} = 226 / (226+13+13+5)$$
$$\text{TP Rate} = 0.879$$

On the other hand, False Positive Rate (FP Rate) is the ratio of the values misclassified in proportion to the total of condition negative.

$$\text{FP Rate} = \Sigma \text{ False positive} / \Sigma \text{ Condition negative}$$
$$\text{FP Rate} = (11+22+2) / (1015-257)$$
$$\text{FP Rate} = 0.046$$

3.3.4.3. Kappa statistic

The kappa coefficient helps us know how good it is a classifier. In the case that the coefficient kappa equals 0, our classifier is behaving like a random classifier. On the other hand, if the coefficient has a lower interval is behaving it will be worse than a random method. As the classifier approaches 1, the classifier is better.

The formula for kappa is:

$$\text{kappa} = (Pr(a) - Pr(e)) / (1 - Pr(e))$$

In our case, Pr(a) indicates the probability that the algorithm guesses right, on the other hand, Pr(e) denotes the probability is guesses by chance.

Example based on the above case:

$$Pr(a) = \text{Correctly Classified Instances} / \text{All instances}$$
$$Pr(a) = 0.841$$

The probability that by chance is classified the plant p2 is given by the probability that it is classified and the probability that it is the sample of the 2nd floor. It is calculated as follows:

$$Pr(e\_p2) = Pr(classifyP2) * Pr(p2)$$
$$Pr(classifyP2) = (226+11+22+2+261)/115=0.257$$
$$Pr(p2) = (226+13+13+5) / 115$$
$$Pr(p2) = 0.253$$

$$Pr(e\_p2) = Pr(classifyP2) * Pr(p2)$$
$$Pr(e\_p2) = 0.065$$

The same with the other plants:

$$Pr(e) = Pr(e\_p2) + Pr(e\_p3) + Pr(e\_p1) + Pr(e\_p0) + Pr(e\_p2)$$
$$Pr(e) = 0.250$$

Finally, kappa coefficient is calculated:

$$kappa = (Pr(a) - Pr(e)) / (1 - Pr(e))$$
$$kappa = (0.841 - 0.250) / (1 - 0.250)$$
$$kappa = 0.788$$

A value greater than 0 means that your classifier is doing better than chance. In the case that the coefficient value is higher, there is a table indicating the quality of the result. In our case, the coefficient is 0.788. So, we can say that the Level of agreement is Substantial.

Table 3.2 Level of agreement kappa coefficient **[36]**

Coefficient	**Level of agreement**
0.00	Poor
0.01 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

3.3.4.4. Recall

Recall is defined as the number of true positives divided by the total number of elements that belong to the positive class. If we look at the first row of the confusion matrix:

$$Recall = 226 / (226+13+13+5)$$
$$Recall = 0.879$$

This measure gives us an idea of how relevant items have been selected.

3.3.4.5. Precision

Precision is defined as the number of true positives divided by the total number of elements that are classified as positive class. If we look at the first row of the confusion matrix:

Precision = 226 / (226+11+22+2)
Precision = 0.866

This measure gives us an idea of how selected items are relevant.

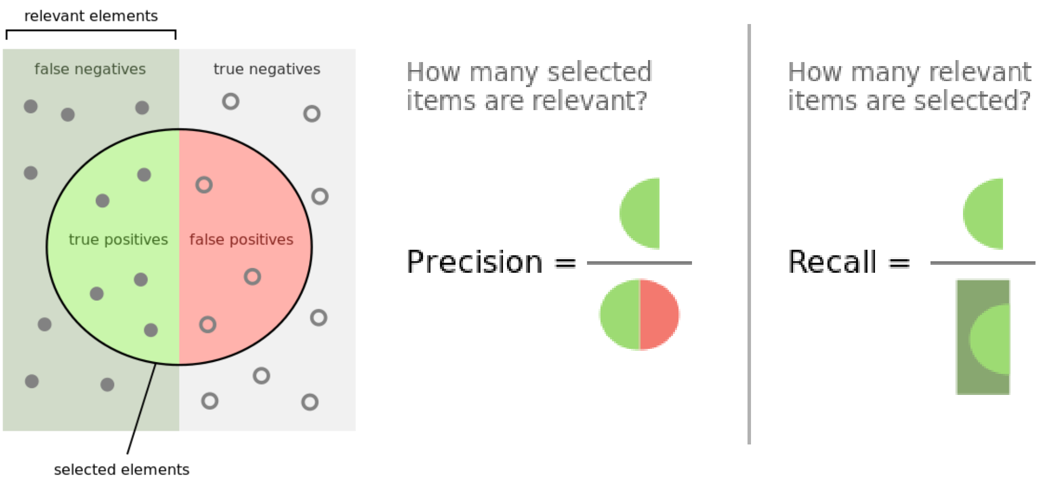


Fig. 3.14 Precision and Recall **[37]**

3.3.4.6. F-measure

F-measure is also named F-score. This value takes into account recall and precision. It is the harmonic mean of precision and recall:

F-measure = 2 * ((precision * recall) / (precision + recall))

If we see the above example:

F-measure = 2 * (0.866 * 0.879 / 0.866 + 0.879)
F-measure = 0.873

3.3.4.7. ROC Area

Receiver Operating Characteristic (ROC) was established to measure the effectiveness of military radars. Later, it was used in medicine and in the statistical analysis. ROC space is defined by FPR and TPR as x and y axes respectively. Each prediction result represents one point in the ROC space. The union of these points draw the ROC curve. The area below this line is called ROC area. The best possible prediction method would yield a point in the coordinate (0,1). It is called a line of non-discrimination to a line that runs along the diagonal from the point (0,0) to (1,1). A completely random guess would give a point along line of non-discrimination. That is why this line marks the boundary between a classification system and a random marking system. A perfect classification system should have an area of roc value 1. In the worst case this value is less than 0.5, which means that our classifier fails more than a random system.

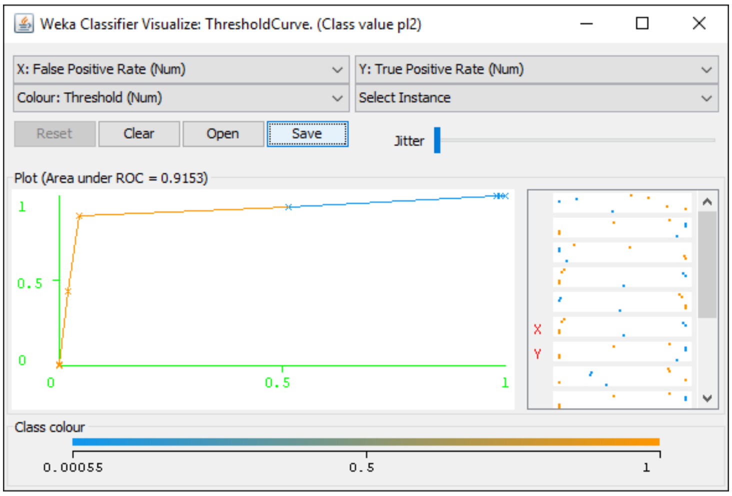


Fig. 3.15 ROC Area graph using WEKA

3.4. Algorithms Analysis

Weka provides a tool called experimenter that allows automated run several algorithms on the same data set. In this section, we will execute several lots of tests to analyze later. We have experimented with the algorithms that are in the following table.

Table 3.3 Algorithms evaluated in this project

Type	**Weka Algorithm**
Naive Bayes	weka.classifiers.bayes.NaiveBaye
Bayes Network	weka.classifiers.bayes.BayesNet
Rule Lerner	weka.classifiers.rules.JRip
Rule Lerner	weka.classifiers.rules.ZeroR
Decision Tree	weka.classifiers.trees.J48
KNN (K- Nearest Neighbors)	weka.classifiers.lazy.IBk
SVM (Support Vector Machine)	weka.classifiers.functions.SMO
Neural network models	weka.classifiers.functions.MultilayerPerceptron

The following figure shows how to define the export route results. The datasets are added and algorithms are selected from the same interface.

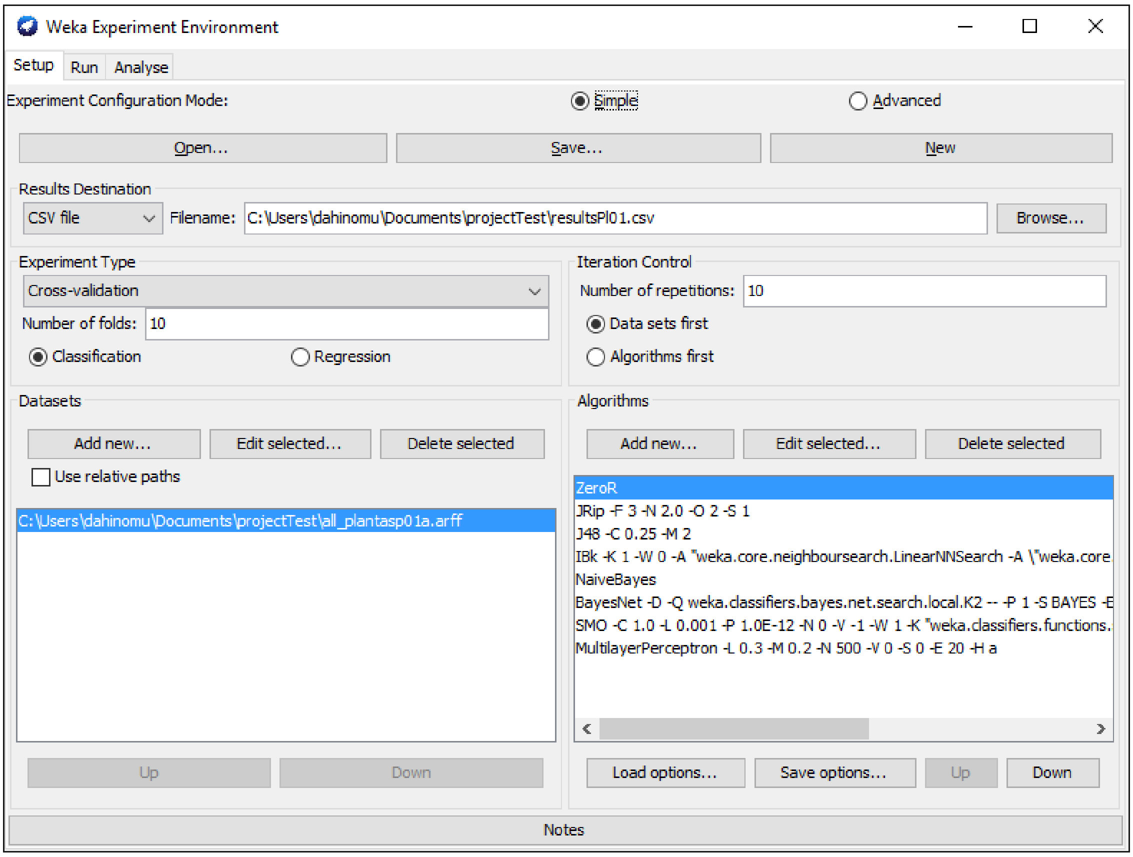


Fig. 3.16 Setting tests in Weka experimenter

The test results are conditioned to the equipment used. In this project, tests have been performed on a computer with the following characteristics:

Table 3.4 Features of testing computer

Features	Value
Processor	i7-5500U CPU @ 2.40GHz
Memory	8GB RAM
OS	Windows 10, 64bit

3.4.1. Success rate versus time response

Key_RUN is an arbitrary parameter between 1 and 10, used to select and divide the database into folders. By default, 10-fold cross-validation divides the database in 10 Folders. Weka pick one to test. Key_Fold defines the number of the folder to be evaluated. We tested all Key_RUN possibilities. In total, there were at least 100 executions for each algorithm. Since we analyzed eight algorithms, we performed 800 executions per database.

In this graph is represented the value of the percentage of success. The columns in the graph show the percentage of success of each test. As may be seen, there is a big difference between success percentage point and the success rate floor.

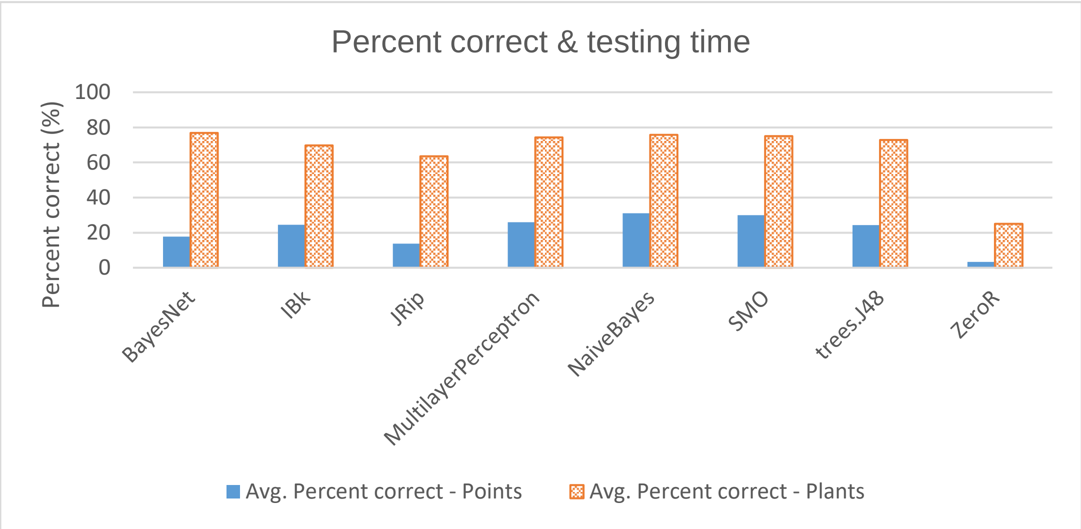


Fig. 3.17 Percent correct depending on the algorithm

On the other hand, the response time is an important factor for choosing an algorithm. The chart below shows the response time of each algorithm. Training time is the time it takes the algorithm to generate a model for prediction. Testing time is the time it takes to classify a sample. To better visualize the results, the graph represents the data logarithmically, although the labels show the response time in seconds. The bigger the bar on the right side, the greater the response time. Conversely, the larger the column on the left side of the axis, the lower the response time. Testing time is an interesting value to measure the quality of service and the resources consumed. Training Time not shown in this figure because it does not affect the quality of service, but does affect the resources consumed during the learning period.

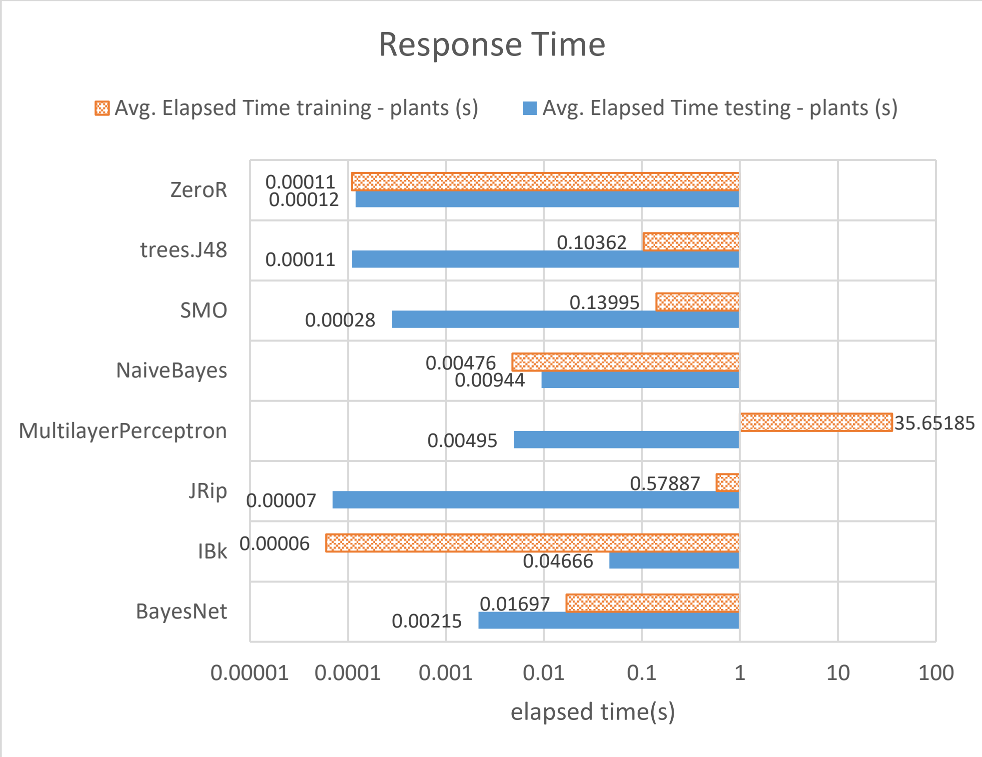


Fig. 3.18 Elapsed Time depending on the algorithm

As noted, ZeroR gives the worst result of success. ZeroR always takes the same value. Therefore, it classifies in the same class all results. It can be considered a random classifier. That's why any classifier that returns a result below this classifier can be discarded from the analysis. Can be seen all the results of other classifiers are better.

Probabilistic classifiers lead the best results, followed closely by the classifier KNN lazy. Unfortunately, the processing time of this algorithm is greater than the rest. This is because the memory access slows the operation. KNN query all data in memory to locate the closest point. We can say that Naive also consumes many resources, but in this case, is due to the complexity of the calculations. It may surprise that the neuronal method consumes so little. Training have been infinitely longer than the other algorithms. Something that directly affects the system administrator. On the other hand, algorithms Rules do not get the best results, but they have a good response time. Finally, we see that Naive Bayes and SMO have good results in both parameters evaluated.

3.4.2. Improving accuracy by splitting the problem

As we saw earlier, the classification results of a percentage point have very low success. In order to improve the results, we have tried to divide the problem into two parts. First the plant is selected. The best algorithms approximate 75% success rate. Then, the point is selected is determined. That should reduce noise from other plants and improves the success rate. Finally, the same test is performed by removing points per plant. The following chart shows the results of this test.

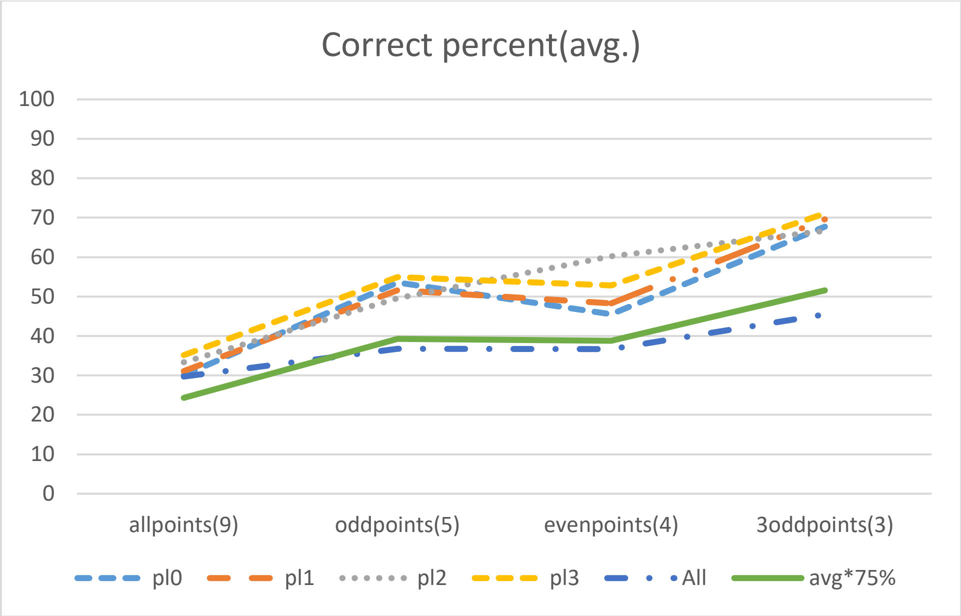


Fig. 3.19 Success rate per floor and the separation between classes

The percentage of success is greater if the floors are separated. But it must be assessed the success rate of the overall system. With the objective of comparing the overall result set with initial values it was performed the following chart. The green line represents the average probability of success after making a preliminary classification of plant. Striped dotted line shows the initial results. It is appreciated that the results of the green line are better as separate points. Unfortunately, this method is only profitable use when the separation point is greater since initially worse result.

3.4.3. ROC Areas

Previously, it has been explained that the area of Roc is for comparing algorithms. In the following chart, Roc area is used to determine which algorithm just gets better result. The graph shows the average of all results in all tests performed. Remember that the value of 0.5 corresponds to a random classifier. That's why ZeroR has a value of 0.5. We see that Naive Bayes has the best average.

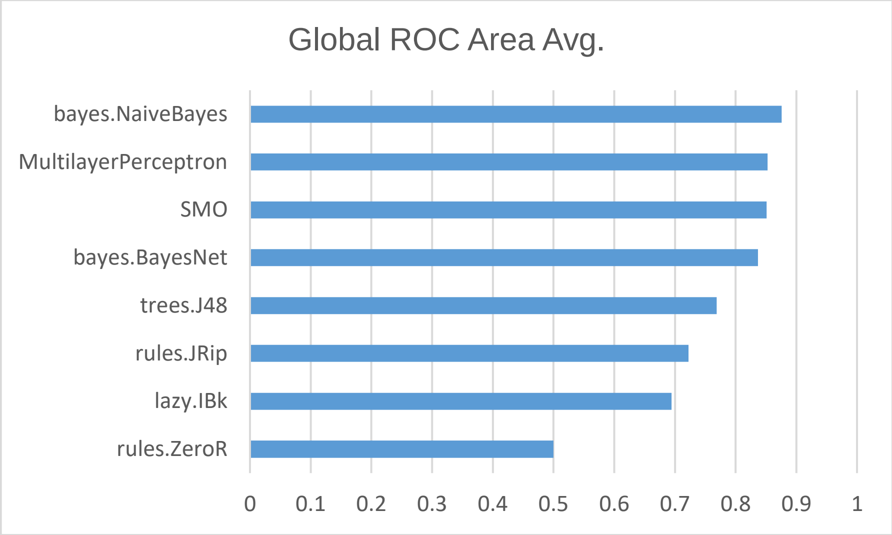


Fig. 3.20 Roc Area according to the algorithm

Table 3.5 Best algorithm ROC Results

ROC	Data selection	Best algorithm	ROC value	ROC lazy.IBk	% success	%succes lazy.Ibk
All plants	Just plants	bayes.BayesNet	0.938	0.800	75.711	69.693
	All points	SMO	0.913	0.616	25.969	13.667
	odd(5)	bayes.NaiveBayes	0.931	0.668	54.775	36.102
	even(4)	bayes.NaiveBayes	0.919	0.674	47.453	38.979
	3points	bayes.NaiveBayes	0.933	0.710	60.227	46.773
plant 0	All points	bayes.NaiveBayes	0.805	0.612	36.720	30.240
	odd(5)	bayes.BayesNet	0.889	0.726	55.967	54.813
	even(4)	bayes.NaiveBayes	0.800	0.621	48.667	43.955
	3points	bayes.NaiveBayes	0.920	0.748	83.393	66.929
plant 1	All points	SMO	0.806	0.599	36.371	28.248
	odd(5)	MultilayerPerceptron	0.854	0.690	55.300	50.433
	even(4)	bayes.NaiveBayes	0.827	0.604	57.309	40.218
	3points	MultilayerPerceptron	0.952	0.784	82.211	71.378
plant 2	All points	SMO	0.832	0.657	45.572	38.957
	odd(5)	bayes.NaiveBayes	0.865	0.687	60.429	50.143
	even(4)	bayes.NaiveBayes	0.870	0.777	69.530	66.833
	3points	MultilayerPerceptron	0.863	0.767	71.819	70.417
plant 3	All points	bayes.NaiveBayes	0.847	0.643	41.843	35.969
	odd(5)	bayes.NaiveBayes	0.917	0.701	69.195	51.710
	even(4)	bayes.NaiveBayes	0.876	0.700	62.515	55.121
	3points	bayes.NaiveBayes	0.955	0.803	83.083	73.319

Tests have been conducted in different scenarios. Initially have been considered all plants, and have been removing the same points on every floor progressively. Then it has done the same floor by floor.

The table 3.5 shows the algorithm with better ROC area in each of the tests. In the case where no points are eliminated, the algorithm overcomes SMO. In other tests, it gives the best result Naive Bayes in most cases. Multilayer Perceptron has also had very good results. In case you want to check all results, see Annex.

3.5. Conclusions of the chapter

This chapter describes all the experimental part of the project and presents the results. In order to carry out this section, 1015 samples were taken distributed over 3 floors of the university campus of Castelldefels. All procedures and tools needed to work with data are shown. One of the main objectives of this section was to showcase a range of possibilities to clean and process data. Another aim was to clarify which algorithm gives better results. That's why thousands of tests were performed. The area of ROC or confusion matrix provide information about the quality of the algorithms. They are really useful tools that help us determine the best algorithm. Without understanding these elements, it would be very difficult to reach accurate conclusions.

Finally, the results are displayed per their success rate, response time and ROC areas. Although my hypothesis was that KNN would deliver better results, it has been shown that its response time is quite low since it must consult data in memory and its success rate is not the best, as it requires a very good workout and it is strongly affected by noise. Probabilistic methods do not consume so much response time and give better results.

CHAPTER 4. CONCLUSIONS

This project is a good starting point for anyone who wants to start in indoor positioning. One of the goals of this project was to present the problem of indoor location. It has tried to be accessible to everyone. The first two topics describe the known technology, mathematics used and current commercial solutions. Throughout, examples are used to make understandable mathematical principles. In the annexes, there is a compilation of companies classified by the technologies. It gives a practical point of view and serves as an example of the dispersity of technologies. This dispersity shows that there is no standard solution to indoor positioning. Furthermore, this project gives the information necessary to understand the classification algorithms most commonly used in machine learning and can serve to anyone who has an interest in the area. This section discusses the results and conclusions of the project.

4.1. Conclusions and results

Traditional positioning is mainly based on triangulation system. The problem is that triangulation ignores multipath, a common problem indoors. There is a method called fingerprinting based on mapping an area and "classification algorithms". In relying on probabilistic and experimental data, the multipath error is reduced. In our case, we have assessed the following algorithms:

- weka.classifiers.bayes.BayesNet
- weka.classifiers.bayes.NaiveBayes
- weka.classifiers.functions.MultilayerPerceptron
- weka.classifiers.functions.SMO
- weka.classifiers.lazy.IBk
- weka.classifiers.rules.JRip
- weka.classifiers.rules.ZeroR
- weka.classifiers.trees.J48

4.1.1. Results

The initial hypothesis was that KNN would better result, but it's easily influenced by noise. We believed that KNN was a good candidate because the first results of our tests were very good. We learned not to use the same data set to train and perform tests.

Table 3.5 compare the results of the best algorithms. This table shows that KNN gives a poorer area of ROC. Neither gave better success rates. It is essential to have a good workout and have a bad response time, as it has to consult all data in memory (as it was shown in Figure 3.18). The following table shows the best algorithms according to the results we have obtained in this work.

Table 4.1 Selection of algorithms according to their use

Class Size	Best Algorithm
Floor	NaiveBayes
Room	NaiveBayes
Position	SMO

Although algorithms Naive and MultiPerception give better results, I conclude that the algorithm SMO is a compromise between response time, resource conservation during training and a good result of success.

4.1.2. Algorithm Conclusions from the results

Below, we present the conclusions we have reached after reviewing the results:

- To determine the position of a user with the accuracy of a point between plants, SMO is the best candidate.
- To determine the position of a user with the accuracy of room or floor, Naive Bayes gives better results.
- For indoor positioning, probabilistic methods are more recommended than others. Their response time is less than the algorithms that query in memory, as KNN and consume far fewer resources than neural networks.
- Neural networks give very good results and its response time is low, but consume a lot of resources in training. The proof of this algorithm was performed independently due to consumption of resources. Otherwise, the computer was blocked because the test was consuming too many resources.
- Methods such as KNN have a bad response time because they have to consult the entire database.
- Methods rules do not give good results. They require a deeper analysis scenario.
- The modification of the number of neighbors KNN algorithm does not help to improve results. Moreover, consider the distance between the neighbor and the sample being classified, it helps improve the success rate.

4.1.3. General conclusions

We believe that we have fulfilled the main objectives of this project:

- It mentioned the main positioning techniques. It has presented the context and indoor positioning technologies. It has been shown how classification tools can help solve the problem. It seemed important to add a list of companies working in the sector.
- From the first moment, actual samples were collected by ourselves to have a comprehensive understanding of the process of working with data. That's why we make a collection and analysis of real samples. It was decided to take 1015 samples at certain points of different plants on the campus of Castelldefels.
- It was used and learned the use of machine learning tools. In our case, we used Weka but also we tried sklearn (python libraries).
- Data preprocessing techniques were learned. In the project, different types of treatments, such as filters and data cleansing are mentioned. We have

conducted thousands of tests. It has deepened in the analysis of ROC area, the response time and the separation between sampling points.

- We have compared 8 different classification algorithms. We have taught how the main algorithms work. To this end, we have described with numerical examples. Thus, one is able to differentiate from each other and pick the one that is needed.

4.1.4. Data conclusions

- The training data collection is one of the most delicate and important parts to work with data.
- The data processing consumes 80% of the time data evaluation.
- Algorithms to select attributes help reduce evaluation costs and help select the most important attributes for the system.
- Never evaluate an algorithm with the same training dataset.
- Data science is an experimental discipline, and that means testing and experiments. Any hypothesis remains a hypothesis until someone shows it experimentally.

4.2. Future work lines

There is still no technology that solves the problem completely. The best solution is to combine different technologies, as appropriate. I would combine the technologies until the devices agree to standardize a method that does not require the introduction of new structures. We must also consider the integration of new technologies in mobile phones that could improve positioning results. As we saw in the first chapter UWB seems to be integrated soon in the new devices. Until that happens I propose the following work steps.

In terms of future improvements, is desirable to perform a functional prototype of this system. An app would consult the data to a server with the maps made. Using the algorithm would return the location that the app would paint on a map. Later, you could add the functionality for guiding. On the other hand, it could be done one business planning.

Once the prototype had been finalized, it should test its performance at the University. For the use of new students in the task of finding the classroom or for locating books in libraries. It could be original use these systems for social events of the university. After checking the proper functioning of the application, it could be known on a larger scale. We can say that this type of application, linked to collaborative game are much in demand.

Taking advantage of the proximity of Mobile World Capital, I think valuable perform a proof of concept in Congress. Such applications have much value in fairs and exhibitions and I think the best place to expose this application would be the mobile Word Congress. Such applications should be open to users but payment for commercial use. I think positive that the university benefits from hosting talent and knowledge it generates.

4.3. Environmental aspects

As explained in this Master Thesis, the proposal of this project has been to use the WLAN standard, existing in most interiors. In this way, it dispenses enable a new infrastructure. The terminal is reused as the signal source. This reduces the economic costs and environmental costs. By using this method, the creation of new specialized devices is not necessary and that reduces the consumption of materials and energy for its creation, unlike other technologies mentioned above.

CHAPTER 5. Bibliography

[1] National Coordination Office for Space-Based Positioning, Navigation, and Timing, "GPS.gov: GPS overview," U.S. Air Force , 11 2 2014. [Online]. Available: http://www.gps.gov/systems/gps/. [Accessed 9 8 2016].

[2] Gibbons Media ;Research LLC, "Inside GNSS," 2009. [Online]. Available: http://www.insidegnss.com/gnssworld. [Accessed 3 5 2016].

[3] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews,* vol. 37, no. 6, p. 1067–1080, 2007.

[4] WiGLE, "WiGLE: Wireless Network Mapping.," 2003. [Online]. Available: https://wigle.net/. [Accessed 3 5 2016].

[5] I. Ahriz, Y. Oussar, B. Denby and G. Dreyfus, "International Journal of Navigation and Observation," *International Journal of Navigation and Observation,* vol. 2010, no. 497829, p. 7, 2010.

[6] decawave, "Decawave and partner Quantitec featured in Nokia booth at 2016 mobile world congress," 2016. [Online]. Available: http://www.decawave.com/decawave-and-partner-quantitec-featured-nokia-booth-2016-mobile-world-congress. [Accessed 9 8 2016].

[7] Philips Lighting Holding, "LED based indoor positioning system | Philips lighting," 2016. [Online]. Available: http://www.lighting.philips.com/main/systems/themes/led-based-indoor-positioning.html. [Accessed 9 8 2016].

[8] H.-S. Kim, D.-R. Kim, S.-H. Yang, Y.-H. Son and S.-K. Han, "An Indoor Visible Light Communication Positioning System Using a RF Carrier Allocation Technique," *Journal of Lightwave Technology,* vol. 31, no. 1, pp. 134-144, 2013.

[9] Accuware, "Visual positioning system (VPS)," 2016. [Online]. Available: https://www.accuware.com/products/visual-positioning-indoors/. [Accessed 9 8 2016].

[10] H. Schantz, "NFER® technology," 30 8 2010. [Online]. Available: http://q-track.com/2010/08/location-innovation-2/. [Accessed 9 8 2016].

[11] J. Liu, "Survey of wireless based indoor localization technologies," 30 4 2014. [Online]. Available: http://www.cse.wustl.edu/~jain/cse574-14/ftp/indoor/. [Accessed 9 8 2016].

[12] The openBmap team, "Wireless collaborative map," 2014. [Online]. Available: http://www.openbmap.org/. [Accessed 9 8 2016].

[13] Geomena, "Geomena," ProgrammableWeb, 28 9 2010. [Online]. Available: http://www.programmableweb.com/api/geomena. [Accessed 9 8 2016].

[14] C. McMullan, D. Cahill and D. Fry, "PlaceLab Geopositioning system," [Online]. Available: http://ntrg.cs.tcd.ie/undergrad/4ba2.05/group1/. [Accessed 9 8 2016].

[15] Skyhook Wireless, "Skyhook | global location and context software products," 2016. [Online]. Available: http://www.skyhookwireless.com/. [Accessed 9 8 2016].

[16] Navizon Inc, "WiFi and Cell-ID location database with global coverage," 2015. [Online]. Available: https://www.navizon.com/. [Accessed 9 8 2016].

[17] K. Lischka, "Google-Debatte: Datenschützer kritisieren W-Lan-Kartografie - SPIEGEL ONLINE," SPIEGEL ONLINE, 22 4 2010. [Online]. Available: http://www.spiegel.de/netzwelt/web/google-debatte-datenschuetzer-kritisieren-w-lan-kartografie-a-690600.html. [Accessed 9 8 2016].

[18] M. B. Kjærgaard, "A Taxonomy for Radio Location Fingerprinting," *Lecture Notes in Computer Science,* p. 139–156, 2007.

[19] Weka, "Class NaiveBayes," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html. [Accessed 9 8 2016].

[20] Weka, "Class OneR," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/OneR.html. [Accessed 9 8 2016].

[21] Weka, "Class JRip," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/JRip.html. [Accessed 9 8 2016].

[22] Weka, "Class J48," 16 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html. [Accessed 9 8 2016].

[23] Wikimedia Commons, "File: Decision tree model.png - Wikimedia commons," 8 10 2005. [Online]. Available: https://commons.wikimedia.org/wiki/File:Decision_tree_model.png. [Accessed 9 8 2016].

[24] Weka, "Class IBk," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/lazy/IBk.html. [Accessed 9 8 2016].

[25] Weka, "Class LibSVM," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.stable/weka/classifiers/functions/LibSVM.html. [Accessed 9 8 2016].

[26] Weka, "Class SMO," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMO.html. [Accessed 9 8 2016].

[27] Wikimedia Commons, "File: Svm max sep hyperplane with margin.png - Wikimedia commons," 16 2 2008. [Online]. Available: https://commons.wikimedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png. [Accessed 9 8 2016].

[28] Wikimedia Commons, "File: Neural network.svg - Wikimedia commons," 28 11 2006. [Online]. Available: https://commons.wikimedia.org/wiki/File:Neural_network.svg. [Accessed 9 8 2016].

[29] Weka, "Class MultilayerPerceptron," 14 4 2016. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/MultilayerPerceptron.html. [Accessed 9 8 2016].

[30] Wikipedia, "Media access control," 7 7 2016. [Online]. Available: https://en.wikipedia.org/wiki/Media_access_control. [Accessed 9 8 2016].

[31] V. Beal, "What is service set Identifider (SSID)? Webopedia definition," 2016. [Online]. Available: http://www.webopedia.com/TERM/S/SSID.html. [Accessed 9 8 2016].

[32] Wikipedia, "MAC address," Wikimedia Foundation, 8 8 2016. [Online]. Available: https://en.wikipedia.org/wiki/MAC_address. [Accessed 9 8 2016].

[33] Weka, "Weka 3 - data mining with open source machine learning software in java," [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/. [Accessed 9 8 2016].

[34] WekaMOOC, "Data mining with Weka," WekaMOOC, 8 9 2013. [Online]. Available: https://www.youtube.com/watch?v=Exe4Dc8FmiM. [Accessed 9 8 2016].

[35] Continuum Analytics, "Download anaconda now!," Continuum Analytics, 22 7 2016. [Online]. Available: https://www.continuum.io/downloads. [Accessed 9 8 2016].

[36] L. J and K. G, "The measurement of observer agreement for categorical data," *Biometrics* , vol. 33, pp. 74-159, 1977.

[37] Wikimedia Commons, "File: Precisionrecall.svg," Wikimedia Foundation, 22 11 2014. [Online]. Available: https://en.wikipedia.org/wiki/File:Precisionrecall.svg. [Accessed 12 8 2016].

[38] Wikimedia Commons, "File: Linear regression.svg - Wikimedia commons," 5 11 2010. [Online]. Available: https://commons.wikimedia.org/wiki/File%3ALinear_regression.svg. [Accessed 9 8 2016].

[39] Wikimedia Commons, "File: Logistic-curve.svg - Wikimedia commons," 2 7 2008. [Online]. Available: https://commons.wikimedia.org/wiki/File:Logistic-curve.svg. [Accessed 9 8 2016].

[40] Wikimedia Commons, "File: KnnClassification.svg - Wikimedia commons," 28 5 2007. [Online]. Available: https://commons.wikimedia.org/wiki/File:KnnClassification.svg. [Accessed 9 8 2016].

[41] Wikimedia Commons, "File: Kmeans toomany.PNG - Wikimedia commons," 21 4 2015. [Online]. Available: https://commons.wikimedia.org/wiki/File:Kmeans_toomany.PNG. [Accessed 9 8 2016].

[42] Observatorio eCommerce de Foro de Economía Digital , "¿Cuál es la cuota de penetración de internet en el mundo?," Observatorio eCommerce de Foro de Economía Digital , 28 7 2015. [Online]. Available: http://observatorioecommerce.com/penetracion-de-internet/. [Accessed 10 8 2016].

5.4. Figures

5.5. Tables

CHAPTER 6. Annexes

6.6. Python code

This code is used to delete or merge points in the database.

```
'''
Created on 5/4/2015
@author: dergal
Listo y leo files de una carpeta.
Tomos los datos de la carpeta y creo una matriz que despues exporto
'''
#!/usr/bin/env python
# -*- coding: utf-8 -*-




def get_algoritm_list(file_path):
    read_data=''
    list=['Scheme','Test mode','Correctly Classified
Instances','Incorrectly Classified Instances','Kappa statistic','Mean
absolute error','Root mean squared error','Relative absolute
error','Root relative squared error','Total Number of
Instances','Weighted Avg.']
    for line in open(file_path):
        for item in list:
            if item in line:
                read_data=read_data+line
    print read_data
    data= read_data.split("\n")
    scheme= data[0].split(":")[1]
    test_mode= data[1].split(":")[1]
    print scheme
    print test_mode
    correctly=data[2].split()[4]
    print correctly
    incorrectly=data[3].split()[4]
    print incorrectly
    kappa=data[4].split()[2]
    print kappa
    Mean_absolute_error = data[5].split()[3]
    print Mean_absolute_error
    Root_mean_squared_error=data[6].split()[4]
    Relative_absolute_error= data[7].split()[3]
    Root_relative_squared_error= data[8].split()[4]
    print Root_relative_squared_error
    #TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area
    print data[10]
    TP_Rate=data[10].split()[2]
    FP_Rate=data[10].split()[3]
    Precision=data[10].split()[4]
    Recall=data[10].split()[5]
    F_Measure=data[10].split()[6]
    ROC_Area=data[10].split()[7]
```

```
algoritm_list=[scheme,test_mode,correctly,incorrectly,kappa,Mean_abso
lute_error,Root_mean_squared_error,Relative_absolute_error,Root_relat
ive_squared_error,TP_Rate,FP_Rate,Precision,Recall,F_Measure,ROC_Area
]
    return algoritm_list

    #Scheme:weka.classifiers.bayes.NaiveBayes
    #Test mode:10-fold cross-validation
    '''
    Correctly Classified Instances         318               31.33
%
    Incorrectly Classified Instances       697               68.67
%
    Kappa statistic                          0.2969
    Mean absolute error                      0.0306
    Root mean squared error                  0.1683
    Relative absolute error                 70.6203 %
    Root relative squared error            114.3079 %

    Total Number of Instances             1015
                   TP Rate   FP Rate   Precision   Recall  F-Measure
ROC Area  Class
    Weighted Avg.    0.313     0.016     0.315     0.313     0.292
0.907
    '''

def get_list_files(path):
    #listar files
    from os import listdir
    from os.path import isfile, join
    onlyfiles = [ f for f in listdir(path) if isfile(join(path,f)) ]
    if 'desktop.ini' in onlyfiles:
        onlyfiles.pop(onlyfiles.index('desktop.ini'))
    return onlyfiles

def export_csv(export_file,matrix):
    import csv;
    '''Export csv'''
    '''export_file_name='exported_nodes.csv'
    export_file=path+export_file_name'''

    with open(export_file, 'wb') as csvfile:
        writer = csv.writer(csvfile, delimiter=';',
                                quotechar='|',
quoting=csv.QUOTE_MINIMAL)
        for row in matrix:
            writer.writerow(row)



def eliminar_puntos(puntos_eliminar,doc,doc_out):

    doc= 'all_plantasR+L.arff'

    f=open(doc,'r')
    file=f.read()
    f.close()
    #print text
    text= file.split('\n')
    print len(text)
```

```python
    print text[98]


#puntos_eliminar=['p02','p04','p06','p08','p10','p12','p14','p16','p0
3','p07','p11','pl0','pl2','pl3','pl4','pl5']
    #puntos_eliminar=['pl0','pl2','pl3','pl4','pl5']

    text_out=[]
    for linea in text:
        if '@attribute location ' in linea:
            #print linea
            inicio=linea.find('{')
            fin=linea.find('}')
            lineaitems=linea[inicio+1:fin]
            linealist=lineaitems.split(',')
            #print linealist


            lista_eliminados=[]
            for punto in puntos_eliminar:
                for item in linealist:
                    if punto in item:
                        print item
                        lista_eliminados.append(item)
            items_out=[]
            for item in linealist:
                if item not in lista_eliminados:
                    items_out.append(item)

            print lista_eliminados
            print items_out



                #linealist=items_out
            #print items_out

            linea_final='@attribute location {'
            for item in items_out:
                linea_final=linea_final+item.replace('"','')+',';
            linea_final= linea_final[0:len(linea_final)-1]+'}';
            print linea
            print linea_final
            text_out.append(linea_final)

                    #index_item=linea.find(item)
                    #linea.find('')
        else:
            punto_detectado = [x for x in puntos_eliminar if x in
linea]
            if len(punto_detectado)==0:
                text_out.append(linea)
    #print text_out

    texto_final=''
    for linea in text_out:
        texto_final=texto_final+linea+"\n"

    print texto_final
    f_out=open(doc_out,'w')
    f_out.write(texto_final)
```

```
    f_out.close()

def main():
    #Planta0 normal
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl0.arff'
    puntos_eliminar=['pl1','pl2','pl3','pl4','pl5']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta0 impares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl0_impares.arff'

puntos_eliminar=['pl1','pl2','pl3','pl4','pl5','p02','p04','p06','p08
']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta0 pares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl0_pares.arff'

puntos_eliminar=['pl1','pl2','pl3','pl4','pl5','p01','p03','p05','p07
','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta0 impares3 saltos
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl0_impares3.arff'

puntos_eliminar=['pl1','pl2','pl3','pl4','pl5','p02','p04','p06','p08
','p03','p07']
    eliminar_puntos(puntos_eliminar,doc,doc_out)

    #Planta1 normal
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl1.arff'
    puntos_eliminar=['pl0','pl2','pl3','pl4','pl5']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta1 impares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl1_impares.arff'

puntos_eliminar=['pl0','pl2','pl3','pl4','pl5','p02','p04','p06','p08
']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta1 pares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl1_pares.arff'

puntos_eliminar=['pl0','pl2','pl3','pl4','pl5','pl5','p01','p03','p05
','p07','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta1 impares3
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl1_impares3.arff'

puntos_eliminar=['pl0','pl2','pl3','pl4','pl5','p02','p04','p06','p08
','p03','p07']
    eliminar_puntos(puntos_eliminar,doc,doc_out)

    #Planta2 normal
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl2.arff'
    puntos_eliminar=['pl0','pl1','pl3','pl4','pl5']
```

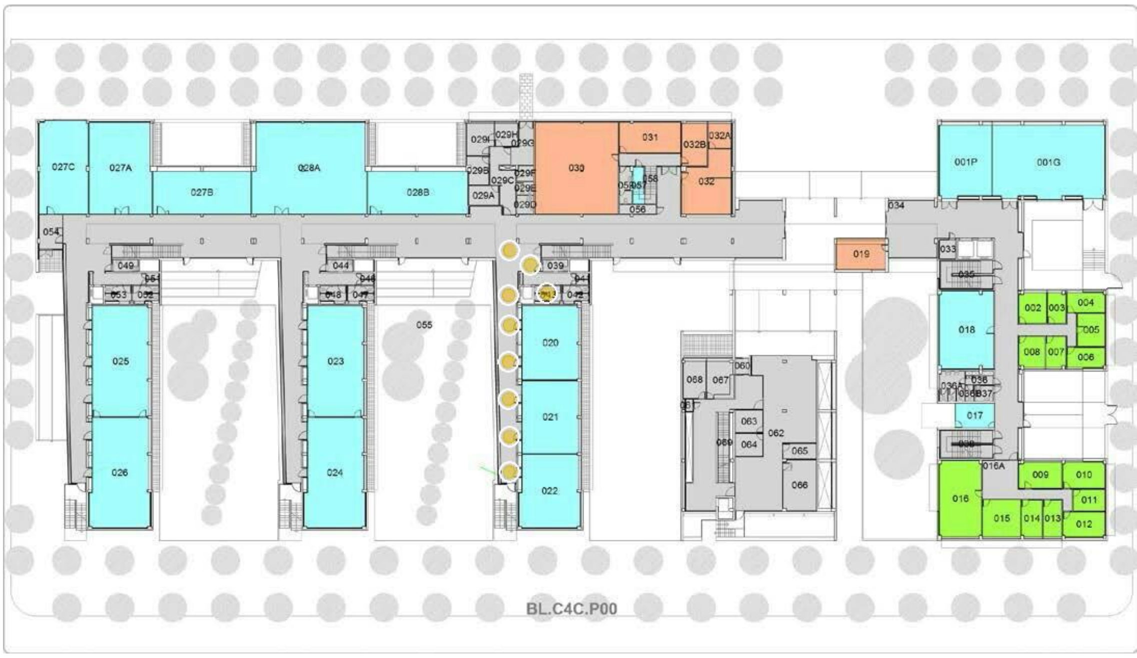
```
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta2 impares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl2_impares.arff'

puntos_eliminar=['pl0','pl1','pl3','pl4','pl5','p02','p04','p06','p08
']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta2 pares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl2_pares.arff'

puntos_eliminar=['pl0','pl1','pl3','pl4','pl5','p01','p03','p05','p07
','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta2 impares3
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl2_impares3.arff'

puntos_eliminar=['pl0','pl1','pl3','pl4','pl5','p02','p04','p06','p08
','p03','p07']
    eliminar_puntos(puntos_eliminar,doc,doc_out)

    #Planta3 normal
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3.arff'
    puntos_eliminar=['pl0','pl2','pl1','pl4','pl5']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p02','p04','p06','p08
']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 pares
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_pares.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p01','p03','p05','p07
','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares3
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares3.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p02','p04','p06','p08
','p03','p07']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares2puntos
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares2puntos.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p02','p04','p06','p08
','p03','p07','p05']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares2puntosjuntos
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares2puntosjuntos.arff'
```

```
puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p05','p04','p06','p08
','p03','p07','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares2puntosseguidos
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares2puntosseguidos.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p02','p04','p06','p08
','p03','p07','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)
    #Planta3 impares3puntosjuntos
    doc= 'all_plantasR+L.arff'
    doc_out='file_out_pl3_impares3puntosjuntos.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p04','p05','p06','p07
','p08','p09']
    eliminar_puntos(puntos_eliminar,doc,doc_out)




def joinPlants(puntos_juntar,doc,doc_out):
    """
    En este algoritmo juntamos todos los puntos por planta
    """
    doc= 'all_plantasR+L.arff'

    f=open(doc,'r')
    file=f.read()
    f.close()
    #print text
    text= file.split('\n')
    #print len(text)
    #print text[98]


#puntos_juntar=['p02','p04','p06','p08','p10','p12','p14','p16','p03'
,'p07','p11','pl0','pl2','pl3','pl4','pl5']
    #puntos_juntar=['pl0','pl2','pl3','pl4','pl5']
    puntos_juntar=['pl0']

    text_out=[]
    for linea in text:
        if '@attribute location ' in linea:
            #print linea
            inicio=linea.find('{')
            fin=linea.find('}')
            lineaitems=linea[inicio+1:fin]
            linealist=lineaitems.split(',')
            #print linealist


            lista_juntados=[]
            for punto in puntos_juntar:
                for item in linealist:
                    if punto in item:
                        #print item
```

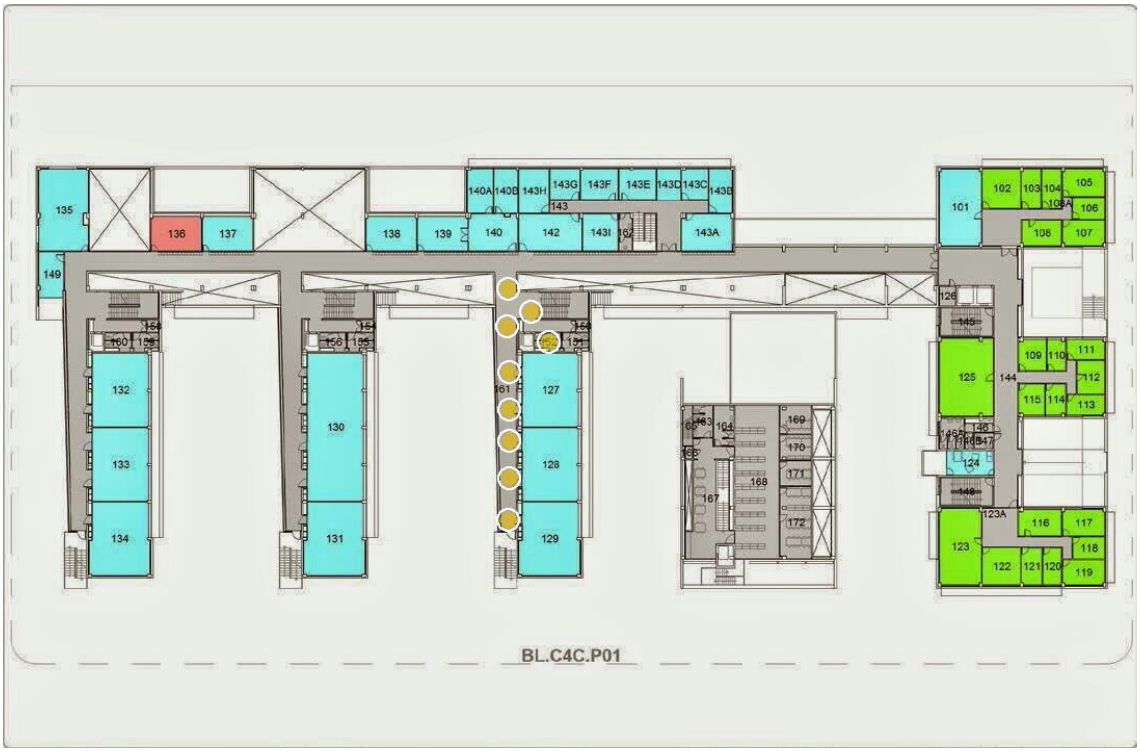
```
                    lista_juntados.append(item)
            items_out=[]
            for item in linealist:
                if item in lista_juntados:
                    print 'myitem'
                    print item
                    items_out.append(item)

            #print lista_juntados
            #print items_out



                #linealist=items_out
            #print items_out

            linea_final='@attribute location {'
            for item in items_out:
                linea_final=linea_final+item.replace('"','')+','
            linea_final= linea_final[0:len(linea_final)-1]+'}'
            #print linea
            print linea_final
            text_out.append(linea_final)

                    #index_item=linea.find(item)
                    #linea.find('')
        else:
            punto_detectado = [x for x in puntos_juntar if x not in
linea]
            if len(punto_detectado)==0:
                text_out.append(linea)
    print text_out

    texto_final=''
    for linea in text_out:
        texto_final=texto_final+linea+"\n"

    '''print texto_final
    f_out=open(doc_out,'w')
    f_out.write(texto_final)
    f_out.close()'''


def joinPlantsMain():
    path='C:/Users/dahinomu/Documents/projectTest/'
    doc= path+'all_plantasR+L.arff'
    doc_out=path+'plantasjuntas.arff'

puntos_eliminar=['pl0','pl2','pl1','pl4','pl5','p04','p05','p06','p07
','p08','p09']
    joinPlants(puntos_eliminar,doc,doc_out)

#start code:
main()
joinPlantsMain()
print 'finished'
```

6.7. Maps

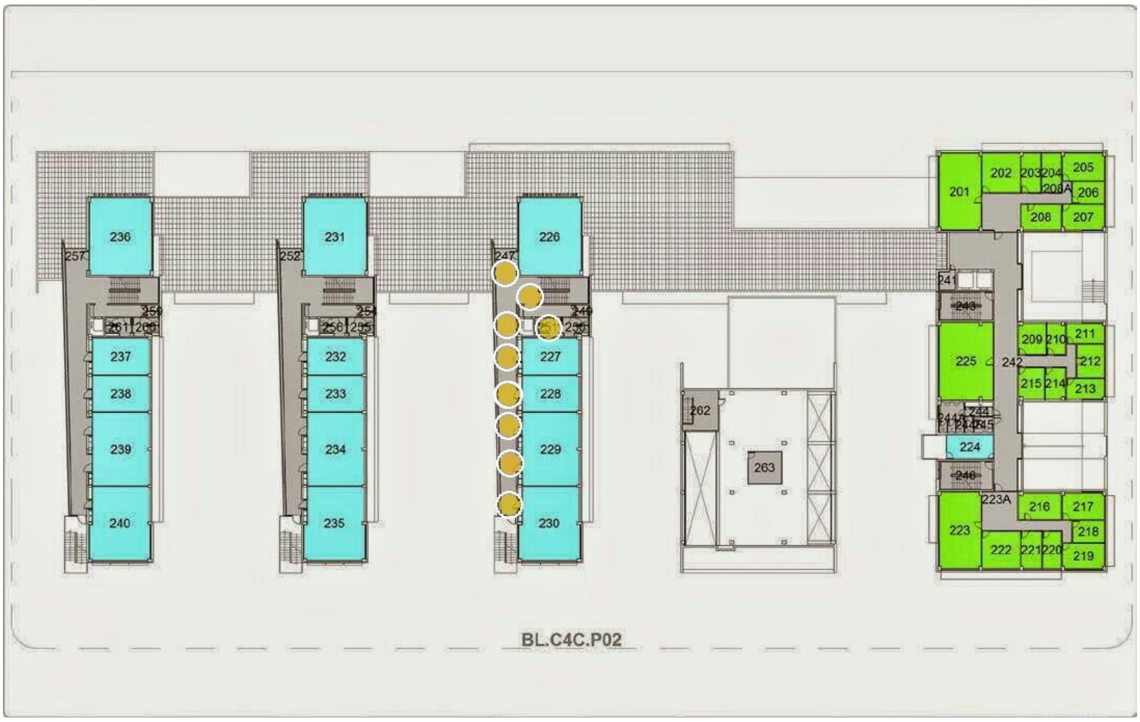
P0



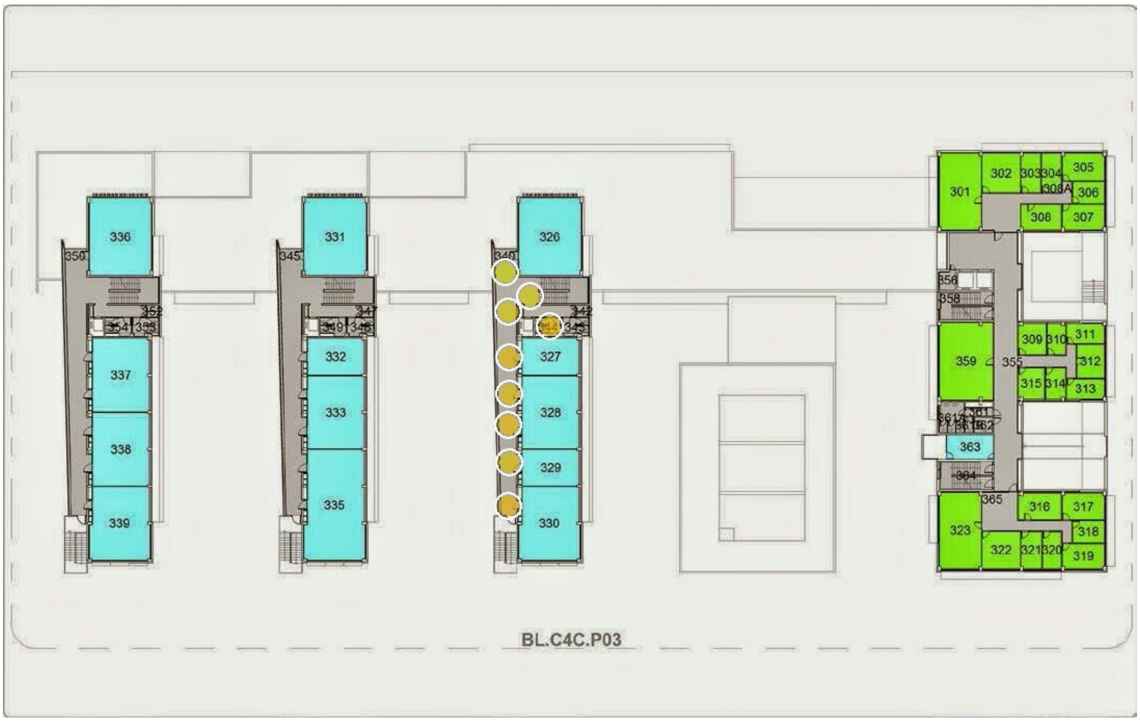
P1



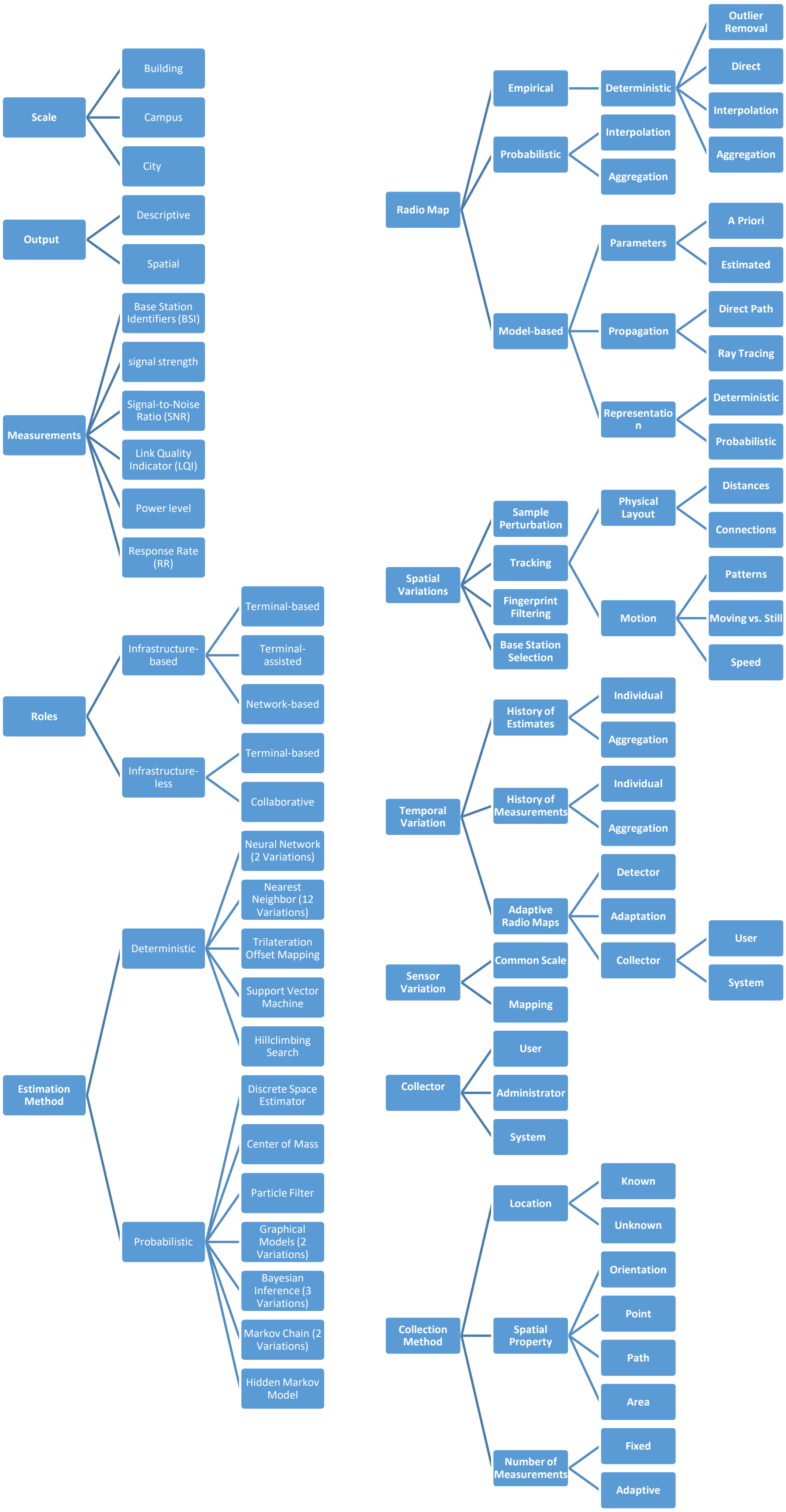
P2


236
231
226
201
202
205
206
207
208
237
238
239
240
232
233
234
235
227
228
229
230
225
209
211
212
213
215
224
262
263
223
216
217
218
219
222
BL.C4C.P02

P3


336
331
326
301
302
305
306
307
308
337
338
339
332
333
335
327
328
329
330
359
309
311
312
313
315
363
323
316
317
318
319
322
BL.C4C.P03

6.8. Taxonomy



Scale
Building
Campus
City
Output
Descriptive
Spatial
Measurements
Base Station Identifiers (BSI)
signal strength
Signal-to-Noise Ratio (SNR)
Link Quality Indicator (LQI)
Power level
Response Rate (RR)
Roles
Infrastructure-based
Terminal-based
Terminal-assisted
Network-based
Infrastructure-less
Terminal-based
Collaborative
Estimation Method
Deterministic
Neural Network (2 Variations)
Nearest Neighbor (12 Variations)
Trilateration Offset Mapping
Support Vector Machine
Hillclimbing Search
Probabilistic
Discrete Space Estimator
Center of Mass
Particle Filter
Graphical Models (2 Variations)
Bayesian Inference (3 Variations)
Markov Chain (2 Variations)
Hidden Markov Model
Radio Map
Empirical
Deterministic
Outlier Removal
Direct
Interpolation
Aggregation
Probabilistic
Interpolation
Aggregation
Model-based
Parameters
A Priori
Estimated
Propagation
Direct Path
Ray Tracing
Representation
Deterministic
Probabilistic
Spatial Variations
Sample Perturbation
Physical Layout
Distances
Connections
Tracking
Motion
Patterns
Moving vs. Still
Speed
Fingerprint Filtering
Base Station Selection
Temporal Variation
History of Estimates
Individual
Aggregation
History of Measurements
Individual
Aggregation
Adaptive Radio Maps
Detector
Adaptation
Collector
User
System
Sensor Variation
Common Scale
Mapping
Collector
User
Administrator
System
Collection Method
Location
Known
Unknown
Spatial Property
Orientation
Point
Path
Area
Number of Measurements
Fixed
Adaptive

6.9. Indoor Companies

We analyzed 47 companies positioning interior.

Name	URL	Description	Country
Tracktio	http://tracktio.com/	Our products – plug&play solutions – helps companies from multiple sectors to track and automate their business operations. Our customers typically need to remotely operate and maintain machines, optimise logistic processes, locate critical assets or people, count available stocks, know their customer's behavior to increase sales, solve operational challenges or avoid logistical errors. All in real time. We can solve your automation and tracking needs with the plug&play solutions we have developed, working together with your trusted system integrator. Partnership is our DNA.	ES
mobius consulting	http://www.mobiusconsulting.com/	independent mobile and wireless consulting services ranging from technology strategy and system analysis to solutions and applications development, by leveraging our extensive experience in WiFi, 4G LTE, mmWave, RFID, RTLS, M2M and IoT. We provide our extensive knowledge and experience in	US

		mobile and wireless technologies, combined with our deep knowledge of target markets, including enterprise, telecom, healthcare, transportation, energy, logistics, and supply chain.	
rf code	http://www.rfcode.com/data-driven-data-center/bid/250586/Active-RFID-Asset-Tags-RTLS-and-the-Internet-of-Things	RF Code Data Center Asset Management and Environmental Monitoring Solutions The data center is at the heart of every thriving company. It doesn't matter whether you own and operate your own data centers; you outsource IT to hosting, cloud or colocation service providers; or you provide hosting/cloud/colocation services to others, the challenge remains the same: maintaining 100% availability while reducing the total cost of running the service. Balancing CAPEX vs OPEX, ensuring regulatory compliance, improving inventorying and auditing processes, and optimizing energy management and capacity planning require absolute accuracy.	US
airista	http://www.airista.com/	AiRISTA develops and manufactures leading Identification & Track and Trace using passive, active, and semi-active RFID; GPS; BLE; and other technologies. We offer	US

		total solutions from RTLS (Real-Time Location Systems) consulting to selecting the right technology components with our platforms in order to provide the best RTLS solution for your business needs. AiRISTA, along with the newly acquired Ekahau RTLS business line of solutions, reveals business intelligence with the power to transform your organization. Use your existing Wi-Fi network (WLAN) to gain real-time visibility into the assets, people, and workflows that drive success with our innovative RTLS platform.	
american rfid solutions	http://www.americanrfidsolutions.net/web/	A significant performance improvement in RFID hardware and tags has led to a leap in capability. ARS' product line provides many opportunities to reduce costs, raise revenue and boost customer satisfaction in the retail sector. Access to more accurate inventory data improves on shelf product availability and reduces inventory costs which leads to higher revenue and profits for both retailers and vendors.The following retail applications can provide significant ROI using RFID and RTLS: Inventory management	US

		– cycle counting Brand authentication Shrinkage – theft and return fraud	
ekahau	http://www.ekahau.com/	Ekahau provides tools for designing, deploying, and maintaining Wi-Fi networks. We're a pioneer in Wi-Fi tools: we made the first ever enterprise-grade site survey tool. Then we built reporting, network planning, and more advanced capabilities.We've done this longer than anyone else. We have over 15,000 happy customers and 500,000 users. Our products are used by Cisco, Aruba, Meru, Aerohive, and the rest of the Wi-Fi folks. And thousands of systems integrators, and IT-administrators.We're a privately owned company. Ekahau investors include Nexit Ventures, Sampo-Mandatum, Tekes – the Finnish Funding Agency for Technology and Innovation and other investors.	Finland
aisle 411	http://www.aisle411.com/	Aisle411 makes it easy to integrate purchase-driving features into retailers' mobile apps including digital in-store maps, product search and location, in-store navigation, and geo-targeted offers. Increase purchase behavior and enhance the shopping experience quickly in three easy steps.	US

		Optimize your store data: Aisle411 creates your searchable digital store and product maps. Integrate: With Aisle411's SDKs, easily integrate maps, search, navigation, and rewards into your mobile apps. Promote: Promote these powerful features to your shoppers and watch guest satisfaction and revenues grow.	
byte light	http://blog.bytelight.com/	The ByteLight™ indoor positioning technology solution from Acuity Brand helps retailers deliver a tailored, mobile-enabled omnichannel shopping experience to their customers. Indoor positioning also can be deployed to optimize and increase understanding of how occupants interact within large-scale indoor environments. These indoor capabilities, together with the ability to provide mapping for surrounding outdoor spaces, create new opportunities for data analytics that can lead to operational efficiencies, enhanced safety, and increased revenues in spaces such as airports, shopping malls, logistics centers, universities and healthcare facilities.	US
zoom shopp er	http://zoomshopper.wix.com/zoomshopper	ZoomShopper is changing the way brands and shoppers interact in the physical world of retail. We aim to create and deliver more	Israel

		engaging and more rewarding experiences for mobile users around the world. Our retail partners benefit from powerful marketing initiatives and campaigns run on our innovative indoor mobile marketing platform. Our platform empowers and rewards mobile users in physical shopping destinations such as big box stores, malls and airports. We help retailers to drive more traffic to their stores and increase their sales. We enable consumer brands and loyalty programs to improve customer loyalty by converting loyalty money into real-time shopping rewards.	
iviutech	http://www.iviutech.com/	Viu Technologies is the market leader in location-aware technologies for interior & exterior foot traffic analytics With the iViu Solution, you can know how your customers shop your store, which will help you maximize their satisfaction and your bottom line success. Learn how iViu can help you bring your customer experience to a new level	US
pozzey	http://www.pozzey.com/	Pozzey is an indoor and micro-location platform for mobile devices such as smartphones and tablets. It provides better than GPS location accuracy to mobile	Australia

		devices both indoors and outdoors using Bluetooth Low Energy beacons (iBeacon) placed throughout a venue to tell the mobile device where it is. Pozzey can provide position accuracy of approximately 3 metres where required. The Pozzey platform provides all the software, maps, databases, beacon hardware, security, analytics and tools that developers need to build feature-rich indoor location enabled apps.	
genius matcher	http://www.geniusmatcher.com/	Map, locate and enjoy! Genius Matcher develops a platform combining computer vision and 3D technology to create one hardware free solution for mapping, positioning and navigating in any indoor location. Our first product for shopping centers, MALLY, with its interactive 3D maps, tells the visitor Where is he standing, What is around him, How to get to each spot and Who is there to share his experience with.	Israel
inloco media	http://www.inlocomedia.com/	MARKET LEADER IN MICRO LOCATION FENCE FOR APP DEVELOPERS No GPS or Bluetooth enabled required. PUSH NOTIFICATION Engage your users in the right place, at the best moment and with the best content. GEO-	US

		TRIGGERED CONTENT Customize your content to deliver information related to your users' context. LOCATION BASED AD-SERVER Run the campaigns you sell directly using the most advanced location targeting capabilities.	
bespoon	http://www.bespoon.com/	BeSpoon is a fabless semiconductor company. We cracked the individual positioning problem. Our chips can track items or individuals within a few centimeters… over long ranges.To achieve that, we measure the time of flight of a ultra wide band signal with a precision of 125 picoseconds. This opens fantastic opportunities to monitor assets, enable precise indoor location and ultimately keep track of all our belongings on our phones.	FR
estimote	http://estimote.com/	Estimote Beacons and Stickers are small wireless sensors that you can attach to any location or object. They broadcast tiny radio signals which your smartphone can receive and interpret, unlocking micro-location and contextual awareness.With the Estimote SDK, apps on your smartphone are able to understand their proximity to nearby locations and objects,	US/Poland

		recognizing their type, ownership, approximate location, temperature and motion. Use this data to build a new generation of magical mobile apps that connect the real world to your smart device.Estimote Beacons are certified Apple iBeacon™ compatible as well as support Eddystone™, an open beacon format from Google.	
meridi an apps	http://meridianapps.com/	Build a better mobile app for your venue. Give your users features that matter: indoor turn-by-turn navigation, indoor location awareness, context-aware information, and more. Navigation: Indoor turn-by-turn directions.Let Meridian digitize your maps. Then provide your visitors with turn-by-turn directions to and from points of interest in your facility.BluDot: Indoor location awareness. Put the 'blue dot' on the map and show visitors where they are in your space. Provide context-aware navigation, or integrate BluDot with Campaigns to send timely info on nearby points of interest.	US
insite o	http://www.insiteo.com/joomla/index.php/en/	Insiteo is a leading company in indoor geolocation and indoor positioning systems (IPS). We propose a complete indoor location platform for the	FR

		development of innovative and high value-added services inside buildings, for different sectors: Retail , Shopping malls , Trade Shows and Exhibitions, Museums, Public Buildings or Offices .With the increase of digital in store and the development of new shopping experiences, location-based services enhance the relevance of mobile applications that you offer to your customers and visitors :Dynamic Mapping, Proximity detection (iBeacon), Accurate Indoor Geolocation (microlocation) Geopush and geofencing, Routing and geolocated shopping lists, Tracking and datamining	
spreo	http://spreo.co/	Indoor Positioning & Navigation Systems SPREO simplifies the process of building location-aware applications. We specialize in providing indoor location software, integration and deployment.Indoor Bluetooth Asset and People Tracking Systems (RTLS) The next generation of Indoor Real-Time Location Systems, powered by Bluetooth. SPREO offers a cost-effective, proven and accurate asset tracking platform. Bluetooth tags	US

		& wearables, cloud & mobile software.	
accuware	https://www.accuware.com/	Accuware is a global provider of indoor location technology. Accuware's products and services have been deployed around the world, powering location-aware applications in many industries, including museums, retail, event marketing, logistics, manufacturing, hospitality, real-estate development, healthcare, casinos and many others.A provider of technology, Accuware relies on a network of Application Partners to deliver industry-specific location-aware applications, and Implementation Partners who build custom systems to clients' specifications.	US
indoo	http://indoo.rs/	This solution is for Shopping centers, individual shops and Supermarkets as well In-store positioning and location-based couponing makes mobile ads much more targeted, relevant and responsive to the customer. Indoor proximity marketing attracts attention of nearby customers and increases the number of visitors. Inside the store, indoo.rs'® technology assists customers in locating and finding	Austria/US

		specific items or make them aware of special offers, which will positively affect their shopping experience and result in the improvement of a retailer‘s business intelligence.Benefits: Spot nearby customers and guide them into your store. Advertise relevant information on products and special offers. Assist customers in finding specific items with in-store navigation. Get insights into customer behavior and feed them back into CRM. Guide a customer through a supermarket according to their shopping list	
polestar	http://www.polestar.eu/en/	Pole Star is a global leader and pioneer in indoor positioning.Created in 2002 by Christian Carle, a former marketing and sales executive at Thales, and Jean Chenebault, a former engineer at Alcatel, Pole Star has ten years proven industry experience and an impressive customer portfolio.NAO Campus®, the Pole Star indoor positioning solution, is based on a unique hybrid technology combining Wi-Fi, GPS, Blue tooth low energy (BLE), MEMS (motion sensors) and map data. Compatible with iPhone and Android, which	Germany/ France/US

		together account for 80% of the Smartphone market, NAO Campus® is the most mature, highest performing indoor positioning service available with over 43 million square feet of indoor coverage worldwide.	
air-go	http://air-go.es/	Air-Go builds Location Based Solutions. Based on our proven, flexible and accurate Indoor Positioning technology we build solutions to enable tracking, guidance, analytics and hyperlocal communication strategies.	ES
lighting	http://www.lighting.philips.com/main/systems/themes/led-based-indoor-positioning.html	Even though the technology's fairly new, it's already hard to imagine life without GPS. Almost everywhere on the planet, people depend on navigation and mobile apps for location-specific information and services. GPS lets people find their way from place to place, get information about their immediate vicinity, and work efficiently.Imagine if you could bring the power of GPS indoors. Retailers could enhance the shopping experience with wayfinding and in-store promotions, giving shoppers new reasons to buy. Businesses could help warehouse staff work more efficiently with easier product finding and	ES

		location-based stocking instructions. Employers could give employees lighting and temperature controls for their workspaces, or could suggest nearby empty meeting rooms based on current location, making them more comfortable and productive.All of this and more is possible today with indoor positioning from Philips	
deca wave	http://www.decawave.com/	DecaWave's DW1000 is the world's first single chip UWB transceiver, enabling you to develop cost effective RTLS solutions with precise indoor and outdoor positioning to within 10 cm. Based on IEEE802.15.4-2011, the DW1000 is also aiming at Internet of Things applications thanks to up to 6.8 Mbps communication capability.	Ireland/France
sewio	http://www.sewio.net/	Sewio solutions use precise location data and wireless sensors to help find, protect, optimize or control the things that matter the most – people, equipment and assets. Using ultra-wide band radio technology, real-time location and modern web based technologies for monitoring and control, Sewio’s solutions are used to protect the safety of employees and machinery, find efficiencies in	Czech Republic

		production and warehousing, and improve processes. Sewio products are used in a wide range of industries including automotive, warehousing and logistics, entertainment, retail, mining and healthcare.	
plus location	http://pluslocation.com/	The PLUS team has developed expertise with real-time location systems starting in 2006, with the design of the first generation of PLUS UWB hardware and tracking software. Since then, PLUS has developed processes for deploying RTLS solutions, with a primary goal of producing the best tracking accuracy and highest reliability. The service that PLUS provides goes far beyond simply deploying a hardware kit. To maintain accuracy and reliability in the system, we have evolved quality processes including understanding in detail the function that the system must provide, performing a survey of the deployment site, simulating the system layout, developing a custom deployment plan, actively managing the installation, and performing qualification testing.	Australia

zebra us	https://www.zebra.com/us/en/solutions/location-solutions.html	Zebra's Visible Value Chain solution for inventory and finished goods tracking provides the detailed information needed by manufacturers, distributors and service companies. This visibility lets you know in real time the arrival, dwell and departure times for every asset throughout the off-line and finished goods processes, and creates a complete history for each asset. See new value for your continuous improvement practices.	ES
ido-link	http://ido-link.com/english/main.action	creating connections that will be mutually beneficial for us and for our customers. “Everything I do, I do it for you!” Under the slogan "Everything I do, I do it for you!", We'll create new history through development of IT integrated intelligent products and solutions with our robust knowledge in technology. First, IDOLINK will pave the way for accompanied growth as a valuable partner of our customers by developing and supplying convenient and tangible products which exceed customers’ expectations. Secondly, IDOLINK will endlessly develop products and services that help our customers’ productivity	Korea

		with our robust knowledge and experience in the field of Internet of Things (IoT) and Internet of Everything (IoE). Lastly, IDOLINK will expand our horizons and do our best to become the global leader in the field of technology through superior technology and global mindsets.	
eliko	http://www.eliko.ee/products/kio-rtls/?gclid=CJiy3-KzrM0CFYEy0wodZ58Ngg	Eliko is an embedded electronics and software development company. Our technologies are found in everything from medical devices to warehouse forklifts. Eliko makes future-proof products and services that are built to last in the changing Internet of Things (IoT) environment. We design technologies based on novel algorithms and communication models for devices that continuously need new functionalities.	Estonia
speed shield	https://www.speedshield.com/index.php/solutions/real-time-location-system-indoor/	The Speedshield RTLS project applies advanced Ultra Wideband (UWB) radio technology in conjunction with MEMS accelerometer sensor fusion to accurately and continuously determine the position of the mobile node in relation to fixed infrastructure and site floorplan references. The technology is small and portable enough to be held on-person, or alternately fitted to	UK/US

		mobile equipment such as forklifts, AGVs and other mobile plant.	
linkpoints	http://linkpoints.co/main.html	Co-ventured with IDOLINK in 2013, LinkPoints has been developing products and services in the field of Internet of Things (IoT) and Internet of Everything (IoE) with its industry leading technology and dedication. Its solution line includes UWB RTLS (Real-Time Location System), OnePass (Smart building system),Smart GAS AMI(Automatic gas meter reading system) which was first commercially applied LPWA system in Korea. LinkPoints is also expanding into B2C space by applying proven technology from various industrial solutions to the consumer products.	US
toportls	http://toportls.com/uwb-rtls/	UWB or Ultra Wide Band is a radio technology. Transmited signals are represented by short impulses. UWB radio frequency range spectrum ranges from 3.1 to 10.6GHz. This range doesn’t let UWB signal to interfere with other, more commonly used by phones and routers conventional frequencies. Like any other radio signal, UWB requers a line of sight(LOS) in order to provide proper accuracy. At the same	US

		time it does not require visual contact with an object. In case of indirect LOS, accuracy would be affected in proportion of the path signal has to travel between two points in comparisson to direct path. In order for UWB system to operate, special tag has to be installed on each object being tracked and several anchors has to be installed in area in which object moves. Because of short messages communication, UWB requires very little enegry to operate.	
indoor atlas	https://www.indooratlas.com/	Indoor positioning systems (IPS) locate people or objects inside a building using radio waves, magnetic fields, acoustic signals or other sensory information collected by a smartphone device or tablet. There are many different types of systems available on the market today including WiFI, Radio and Beacons. IndoorAtlas provides a unique cloud platform that runs a disruptive geomagnetic positioning in its core to accurately pinpoint a location inside a building.	US/Finland
prism	https://prism.com/	Transform your cameras and other connected devices into analytics tools so you can remotely monitor, manage, and optimize	US/UK

		your business. Data-driven Decisions, Made Simple. Prism gives you everything you need to effectively manage and analyze all your locations around the world. VISUAL ACCESS Get anytime, anywhere access to what's happening in every one of your locations. Eliminate the guesswork and optimize your business with actionable insights.	
sonitor	http://www.sonitor.com/	Sonitor Technologies is a leading provider of Real Time Location System (RTLS) solutions and develops, manufactures and supplies a proprietary ultrasound indoor positioning system (IPS) that automatically tracks with 100% room or sub-room accuracy the real-time location of moveable equipment and people in complex indoor environments, such as hospitals. Sonitor Technologies' RTLS technology is designed for seamless integration with third party applications software and integration solutions.	US/Norway
indoor lbs	http://www.indoorlbs.com/#ibeacons-1	In-Store Wayfinding & Product Locator Every customer needs help finding items in your store, and they prefer to use their smartphone for the task! In-Store Wayfinding & Product Locator Every customer needs help finding items	US

		in your store, and they prefer to use their smartphone for the task! We work with you on your requirements, timelines, and budget. Based on this information, we seek out the best solution. IndoorLBS is an independent consulting company that works with retailers, airports, malls, travel and shopping apps and technology providers. We work with over 200 vendors to bring the right solution for you.	
accuware	https://www.accuware.com/	Locate, Track and Monitor people and assets in the real world Our systems enable businesses to make their mobile apps and their venues location-aware, delivering accurate positioning everywhere We provide solutions to your location needs Our software provides accurate location both globally and indoors, making your mobile apps and applications aware of physical location in the real world	US
pointrlabs	http://www.pointrlabs.com/	Meet the best performing indoor location and data platform Pointr is an innovative indoor technology solution that's multiplatform, easy to install and cost effective to maintain. At Pointr, we are industry experts	US/UK/DEU

		in providing a real time blue dot within your venue's map. Our deployments range from positioning users in busy venues such as airports to asset tracking in warehouses. We provide state of the art cloud platforms that help clients easily manage their venues, location based communications and showcase insightful data analytics.	
beaconstac	http://www.beaconstac.com/	Everything you need to get started with your beacon project. Deploy beacons, create proximity marketing campaigns and track meaningful analytics, without writing a single line of code. iBeacon & Eddystone-compatible beacon hardware. Build your project with beacons powered by replaceable AA batteries that last up to 4 years at default settings. Easy-to-use Beaconstac app Manage your fleet of beacons and experience your first proximity-based campaigns using the Beaconstac app.	US
infsoft	http://www.infsoft.com/	Indoor Navigation by infsoft. infsoft Indoor Positioning infsoft Indoor Navigation enables better orientation in complex building structures. Using our technology, the position of a smartphone can be pinpointed to 1 meter	DEU

		(approx. 3 feet) accuracy without installing costly hardware. In the field of location-based marketing and of geo-based offers, this allows for new application scenarios. infsoft solutions are already being used successfully around the world at trade fairs, in airports, in hospitals, in office buildings and in shopping malls. Indoor Navigation by infsoft infsoft Indoor Positioning infsoft Indoor Navigation enables better orientation in complex building structures. Using our technology, the position of a smartphone can be pinpointed to 1 meter (approx. 3 feet) accuracy without installing costly hardware. In the field of location-based marketing and of geo-based offers, this allows for new application scenarios. infsoft solutions are already being used successfully around the world at trade fairs, in airports, in hospitals, in office buildings and in shopping malls.	
air-fi	http://www.air-fi.es/	Needs Companies demand solutions WiFi High Capacity Sufficient bandwidth and density capable of replacing the Structured Ethernet cabling to Workstation	ES

		Users , offering well , the UN Environment flexible and safe work (BYOD) for Movements and changes changing the very nature of the business requires. Sistemas de Localización en Tiempo Real (RTLS). Los sistemas de localización en tiempo real (RTLS) e Indoor Positioning Solutions (IPS) ofrecen a los administradores de red la capacidad de identificar y rastrear la ubicación física de los dispositivos en tiempo real para la gestión de inventarios, el seguimiento de la ubicación de una persona y la optimización de flujos de operativos.	
en.situm	https://en.situm.es/	Situm has developed an accurate and consistent multi-sensorial technology that can be adapted to any space, avoiding the installation of additional hardware. Indoor Maps SDK. Situm Indoor Maps is a software (SDK) for smartphones that allows you to integrate our technology into any app easily or even create a new one from scratch.	ES
mysphera	http://mysphera.com/	Real Time Location Systems are completely automatic systems which monitor the position of a mobile element using a specific frequency. Both real time location systems services (RTLS) and	ES

		identification systems are based on radiofrequency (RFID). The application/implementation of this type of systems are becoming widely accepted in hospital environment due to its high return on investment (ROI), with respect to human and material resources management, increasing the clinical safety of patients. It is important to choose the right system as many systems are not specifically designed for hospital environments and cannot carry out the needs of its clients. More important than the technology used, are the advantages offered by a Real Time Location System for hospitals and their different departments.	
ezentis	http://www.ezentis.com/	The main objective of the project is to provide the product Intellimap with the functionalities necessary, in order to develop a mobility application on mobile devices (PDA), to facilitate and optimize working-forces intervention on the ground. We need to fix as the fundamental objective of this project, that of to make progress and consolidate the GIS technology as a national product and promote it at an international level. A command and control	ES

		system for the Swiss army which uses Intellimap as a GIS solution, can be taken as reference. Intellimap-Mobility aims to respond to maintenance technicians, technical support, security forces, engineers, armies, census workers, etc.., who use or will use GIS mobile technology to accomplish their tasks. The most important interventions that the proposed application will grant:,processing work orders geographical people traking field data collection, technical inspections, Incident Report, maintainance of inventory records, data browsing and editing via Web, 3D visualization. Intellimap-Mobility will integrate one or more of the already existing on market technologies related with Communication (Wireless), Positioning (GPS), Mobility (PDA) and 3D Simulation.	
nebus enses	http://www.nebusens.com/es/	Advantages of Polaris versus other RTLS. Lower total cost compared with systems based on Wi-Fi or RFID, and much lower than those based on UWB. Higher tolerance to the presence of walls and obstacles than systems based on Wi-Fi and UWB. n-Core Sirius B devices provide much longer battery life (even	ES

		months) compared with devices based on Wi-Fi and UWB. Its performance is not affected by Wi-Fi networks, thanks to the greater number of used channels. Frequency band approved for its use in industrial and medical environments. Includes automation features for controlling all kinds of sensors and actuators.Web Services based architecture that facilitates the integration of n-Core Polaris with a wide range of applications, including mobile interfaces. Capability to create your own RTLS by means of the n-Core API (Application Programming Interface).	
Ubisense	http://ubisense.net/en	Ubisense is a global leader in Enterprise Location Intelligence solutions. We work with manufacturing, communications and utilities companies, enabling them to improve operational effectiveness, increase efficiency and boost profitability. With the latest location technology at their fingertips, the possibilities are endless.	US/UK/FR/DEU
Q-track	http://q-track.com/	Q-Track provides Real-Time Location Systems (RTLS) – similar to GPS, but more accurate and capable of working indoors. Q-Track's Near Field Electromagnetic Ranging (NFER®)	US

		products keep workers safe from collisions with automated cranes in dozens of factories nationwide. The nuclear power industry was an early adopter of NFER® products, using them to train employees to avoid exposure to radiation. NFER® products track soldiers in a large training facility for Military Operations Urban Terrain (MOUT). Q-Track's products successfully track oil and gas workers, hockey players, and integrators are exploring other applications.	
Gimbal	http://www.gimbal.com/	A single enterprise-class platform for location and proximity, powerful mobile engagement, analytics and visualization solutions with reach across 100s of millions of devices. See why the world's leading enterprises use Gimbal.	US

6.10. ROC Area results

ROC	Algorithms	bayes.Bayes	bayes.Naive	lazy.IBk	MultilayerPe	rules.JRip	rules.ZeroR	SMO	trees.J48	Best algorithm	value
	Just plants	0.93840011	0.93274501	0.80003365	0.91181028	0.81475583	0.5	0.88636442	0.87658793	**bayes.BayesNet**	**0.93840011**
	All points	0.86426369	0.90945688	0.615953	0.86026913	0.64384091	0.5	0.91336212	0.68657318	**SMO**	**0.91336212**
	odd(5)	0.89028427	0.93138491	0.66844917	0.90361797	0.74808196	0.5	0.9223908	0.77421434	**bayes.NaiveBayes**	**0.93138491**
	even(4)	0.88500061	0.9186342	0.67445369	0.87951658	0.72354295	0.5	0.90027167	0.75943257	**bayes.NaiveBayes**	**0.9186342**
All plants	**3points**	0.89145509	0.93314907	0.71006962	0.92762403	0.77740952	0.5	0.91765212	0.81213284	**bayes.NaiveBayes**	**0.93314907**
	All points	0.78655078	0.80533322	0.6119734	0.76555092	0.65947533	0.5	0.79436232	0.68626896	**bayes.NaiveBayes**	**0.80533322**
	odd(5)	0.88946118	0.88421688	0.72620109	0.8757354	0.71316982	0.5	0.87583707	0.75890874	**bayes.BayesNet**	**0.88946118**
	even(4)	0.73121586	0.79986466	0.62085617	0.76120007	0.69865804	0.5	0.74113119	0.67882595	**bayes.NaiveBayes**	**0.79986466**
plant 0	**3points**	0.83814583	0.92036012	0.74757738	0.90897619	0.69986012	0.5	0.90656548	0.76727381	**bayes.NaiveBayes**	**0.92036012**
	All points	0.76513287	0.79262375	0.5992958	0.78355879	0.67527421	0.5	0.80613203	0.67627351	**SMO**	**0.80613203**
	odd(5)	0.82446275	0.84563989	0.68993298	0.85415809	0.79613109	0.5	0.85159151	0.78215738	**MultilayerPerceptron**	**0.85415809**
	even(4)	0.77287599	0.82710164	0.60371988	0.78009506	0.67260768	0.5	0.81093371	0.75699847	**bayes.NaiveBayes**	**0.82710164**
plant 1	**3points**	0.85526587	0.88284259	0.78366071	0.95171958	0.82963955	0.5	0.87016138	0.84109557	**MultilayerPerceptron**	**0.95171958**
	All points	0.77096938	0.82824079	0.65749194	0.78026252	0.62909173	0.5	0.83176008	0.69842908	**SMO**	**0.83176008**
	odd(5)	0.83312825	0.86472835	0.68650108	0.83878139	0.70654383	0.5	0.85273106	0.80524621	**bayes.NaiveBayes**	**0.86472835**
	even(4)	0.80559411	0.87008063	0.77747817	0.83224125	0.78817002	0.5	0.83519048	0.7729007	**bayes.NaiveBayes**	**0.87008063**
plant 2	**3points**	0.83937963	0.85651091	0.76692444	0.862625	0.73991551	0.5	0.81343469	0.78070453	**MultilayerPerceptron**	**0.862625**
	All points	0.81899639	0.84701413	0.64257655	0.818265	0.68591921	0.5	0.8300923	0.73346359	**bayes.NaiveBayes**	**0.84701413**
	odd(5)	0.87088137	0.91738693	0.70122794	0.86888628	0.74191098	0.5	0.84804845	0.85005961	**bayes.NaiveBayes**	**0.91738693**
	even(4)	0.80626514	0.87569959	0.70014087	0.8182926	0.64951578	0.5	0.81294578	0.75708425	**bayes.NaiveBayes**	**0.87569959**
plant 3	**3points**	0.89301968	0.95475198	0.8032381	0.91842295	0.77705308	0.5	0.85003108	0.88583813	**bayes.NaiveBayes**	**0.95475198**
Global	**Average**	0.83670233	0.8760841	0.69465503	0.85245757	0.72240796	0.5	0.85099951	0.76859378	**bayes.NaiveBayes**	**0.8760841**