



UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA



Localización Indoor Basada en Wi-Fi

MEMORIA DE PROYECTO PRESENTADA A LA FACULTAD DE
INGENIERÍA DE LA UNIVERSIDAD DE LA REPÚBLICA POR

Antonio Bracco, Federico Grunwald, Agustín Navcevich

EN CUMPLIMIENTO PARCIAL DE LOS REQUERIMIENTOS
PARA LA OBTENCIÓN DEL TÍTULO DE
INGENIERO.

TUTOR

Dr.-Ing. Germán Capdehourat Universidad de la República
Dr.-Ing. Federico La Rocca..... Universidad de la República
Dr.-Ing. Eduardo Grampín..... Universidad de la República

TRIBUNAL

M.Sc. Ewelina Bakala Universidad de la República
Dr.-Ing. Pablo Musé Universidad de la República
M.Sc.-Ing. Matías Richart..... Universidad de la República
Prof. María Simon Universidad de la República

Montevideo
miércoles 1 mayo, 2019

Localización Indoor Basada en Wi-Fi, Antonio Bracco, Federico Grunwald, Agustín Navceovich.

Esta tesis fue preparada en L^AT_EX usando la clase iietesis (v1.1).
Contiene un total de 137 páginas.
Compilada el miércoles 1 mayo, 2019.
<http://iie.fing.edu.uy/>

Cuando tenés razón, tenés razón.

PROVERBIO POPULAR

Esta página ha sido intencionalmente dejada en blanco.

Agradecimientos

Este proyecto no hubiera sido posible sin el apoyo y colaboración de una cantidad enorme de personas a las cuales queremos expresarles nuestro agradecimiento.

Primeramente nuestro mayor agradecimiento es hacia Federico y Germán, por el apoyo continuo y diligente, el conocimiento transmitido y el acceso a diversos recursos. Una oración aparte merece el papel mayúsculo que tuvieron en todo lo que tiene que ver con negociaciones, permitiéndonos trabajar en la mejor parte.

Queremos agradecer al Museo Nacional de Artes Visuales (MNAV) por ofrecernos la oportunidad de trabajar en un proyecto tan interesante, por siempre confiar en nosotros y por la buena disposición de todos sus funcionarios involucrados. En especial mención a Enrique Aguerre director del MNAV que, junto a la Decana María Simon, lograron que todo esto fuera posible. A Pablo Uribe por dejarnos entrometer la tecnología en su muestra artística. También, a Maru Vidal y Marina García que nos ayudaron en todo el despliegue de la solución en el museo. Por último, a todo el personal de seguridad que nos acompañó y nos trató de la mejor manera en las largas horas que pasamos en el museo.

A Antel, por apoyarnos en la realización de este proyecto. Sin su apoyo económico y logístico no hubiese sido posible desarrollar el despliegue y realizar las pruebas necesarias para que este proyecto fuera más que una idea. En especial se destaca a Leonardo Kammermann por su labor como nexo con Antel.

A Tryolabs, por facilitarnos el uso de sus instalaciones para trabajar en nuestro proyecto que sin dudas no hubiera sido posible sin el espacio de trabajo que nos brindaron.

A nuestras amigas de Arquitectura que nos facilitaron los planos del museo con el cual pudimos hacer las distintas planificaciones de despliegues en el museo.

Y por último pero no menos importante agradecer a todas nuestras familias, novias y amigos que nos acompañaron en todo el transcurso de este proyecto. Gracias por su constante apoyo a lo largo de toda la carrera, ayudando desde lo más simple hasta colaborando activamente en este proyecto.

Esta página ha sido intencionalmente dejada en blanco.

A toda la gente que nos acompañó en el camino.

Esta página ha sido intencionalmente dejada en blanco.

Resumen

Los sistemas de localización más utilizados en la actualidad están basados en GPS (Global Positioning System), con un rendimiento pobre en ambientes cerrados tales como museos, comercios o aeropuertos. Para este tipo de situaciones existen diversas tecnologías alternativas que permiten a un dispositivo móvil ubicarse o ser ubicado con mayor precisión. En el presente proyecto se repasan las soluciones basadas en Wi-Fi, desarrollando un sistema y una aplicación de Android que hagan uso de esta solución. La aplicación desarrollada para la muestra Aquí soñó Blanes Viale en el MNAV (Museo Nacional de Artes Visuales) y operativa desde el 8 de noviembre de 2018 hasta el 3 de marzo del 2019, proporciona contenido en función de la posición del usuario. Provee adicionalmente una mayor accesibilidad a la muestra para usuarios con discapacidad visual, haciendo especial uso de la tecnología de localización para evitar la necesidad de interacción con la pantalla. La recepción del público fue positiva, fomentando la idea de que este tipo de aplicaciones poseen un gran potencial para las tecnologías de localización indoor. El sistema utiliza 15 APs instalados en el museo, dividiendo el mismo en 16 zonas. Proporciona una precisión promedio de 95,7 % y una buena experiencia de usuario. Posteriormente, se realizó un estudio del desempeño del sistema bajo diversas situaciones y un breve estudio del recorrido de los usuarios que hicieron uso de la aplicación.

Esta página ha sido intencionalmente dejada en blanco.

Tabla de contenidos

Agradecimientos	III
Resumen	VII
1. Introducción	1
1.1. Antecedentes	1
1.1.1. Museo Nacional de Artes Visuales	1
1.1.2. Tecnologías de localización indoor	2
1.2. Objetivos	3
1.3. Alcance	3
1.3.1. Accesibilidad	3
2. Estado del arte	5
2.1. Introducción	5
2.2. Alternativas a WLAN para localización indoor	5
2.2.1. BLE	6
2.2.2. Campo magnético	7
2.2.3. Acelerómetro y giroscopio	8
2.2.4. RFID	8
2.3. Localización basada en WLAN	8
2.3.1. Received Signal Strength Indication: RSSI	9
2.3.2. Beacon frames	9
2.3.3. Escaneo activo y pasivo	9
2.3.4. Trilateración	10
2.3.5. Sistema de huellas digitales	10
2.4. Clasificación y aprendizaje utilizando huellas	12
2.4.1. Algoritmos	12
2.5. Combinación de algoritmos para mejorar exactitud	18
2.5.1. Motivación	18
2.5.2. Índice de Youden	19
2.6. Consideración de la variable temporal	21
2.6.1. Filtro Bayesiano	21
2.6.2. Filtro de Kalman	21
2.7. Elección y descripción de la arquitectura del sistema	22
2.7.1. Framework for Internal Navigation and Discovery: FIND3	23

Tabla de contenidos

2.8. Comentarios sobre el estudio del problema	25
3. Implementación de la solución	27
3.1. Introducción	27
3.2. Pruebas preliminares	27
3.3. Distribución de APs	28
3.3.1. Diseño inicial	28
3.3.2. Modelo de APs utilizados	29
3.3.3. Configuración de APs	30
3.3.4. Instalación final y resultados	31
3.4. Arquitectura del sistema y modificaciones a FIND3	33
3.4.1. Arquitectura general	33
3.4.2. Componentes de la arquitectura	35
3.4.3. Interacciones entre los componentes	36
3.4.4. Aplicación de recolección de huellas	38
3.4.5. Corrección en utilización de huellas	39
3.5. Distribución de zonas de localización	39
3.5.1. Consideraciones realizadas	40
3.5.2. Relevamiento de primeras huellas	40
3.6. Aplicación de Android: MNAVegante	41
3.6.1. Interacción entre aplicación y sistema de localización	42
3.6.2. Información general	44
3.6.3. Exhibición	44
3.6.4. Recorrido a ciegas	46
3.6.5. Aplicación web para información de museo	48
3.7. Inauguración de la muestra	49
4. Análisis de desempeño	51
4.1. Introducción	51
4.2. Metodología y nomenclatura utilizada	52
4.2.1. Criterio de aceptación del sistema	53
4.2.2. Conjuntos de entrenamiento, prueba y validación	54
4.2.3. Omisión de algoritmos	54
4.3. Primer acercamiento al problema	55
4.3.1. Representación en potencia de APs no detectados	57
4.4. Entrenamiento con selección de MACs	58
4.4.1. Filtrado según APs instalados	58
4.4.2. Entrenamiento para dispositivos de una banda	60
4.4.3. Precisión en función de cantidad APs instalados	62
4.4.4. Precisión en función de posibles fallas en APs	66
4.5. Precisión en función de cantidad de muestras	66
4.6. Resolución espacial	68
4.7. Resumen de resultados	71

5. Estadísticas de uso	73
5.1. Introducción	73
5.2. Actividad de usuarios en el tiempo	73
5.3. Zonas más concurridas	75
5.4. Estadísticas de la aplicación	76
6. Conclusiones y líneas de trabajo a futuro	79
6.1. Conclusiones	79
6.1.1. Introducción	79
6.1.2. Experiencia de usuario y recepción del público	79
6.1.3. Análisis de desempeño	80
6.2. Trabajo a futuro	81
6.2.1. Introducción	81
6.2.2. Dimensionamiento en cantidad de APs y huellas	81
6.2.3. Solución a dispositivos sin doble banda	81
6.2.4. Filtro de Kalman	82
A. A.1 Reporte Ekahau	83
Referencias	113
Índice de tablas	116
Índice de figuras	118

Esta página ha sido intencionalmente dejada en blanco.

Capítulo 1

Introducción

1.1. Antecedentes

1.1.1. Museo Nacional de Artes Visuales

En el Museo Nacional de Artes Visuales (MNAV) existe una sección de la exposición con elementos complementarios que permiten que personas de baja visión o no videntes puedan disfrutar de las obras presentes. Este conjunto, denominado Museo Amigo, cuenta con un sistema de lectura braille y tablets, que se prestan a quienes lo soliciten.

Por otro lado, se plantea inaugurar la muestra Aquí Soñó Blanes Viale, proyecto expositivo de Pablo Uribe. Esta muestra, a desarrollarse entre el 8 de noviembre de 2018 y el 3 de marzo de 2019, tomaría como insumos principales el acervo del MNAV y la propia arquitectura del mismo, ocupando todas las salas y espacios comunes del MNAV y realizando cambios estructurales sobre ellos. Entre los cambios estructurales que se realizarían, se removerían los totems que constituyen al Museo Amigo. En búsqueda de mantener la accesibilidad previamente proporcionada por el Museo Amigo, el MNAV consultó a la Facultad de Ingeniería de la Universidad de la República por alternativas que agregaran una mayor independencia a los usuarios.

Es en este contexto que se plantea proveer al MNAV con una aplicación para dispositivos móviles que trabaje con la muestra, garantizando un acceso a contenido complementario a la muestra a medida que el usuario se encuentre próximo a distintas obras. La misma proveerá un servicio adecuado tanto para usuarios con dificultades de visión como para el resto del público general. Esta aplicación se basaría en la implementación de un sistema de localización indoor utilizando señales de Wi-Fi. Se busca así lograr que la experiencia del usuario tenga una interacción con su posición dentro del museo. La elección de esta tecnología se realizó basado en lo que se discutirá en el capítulo 2, pero sobre todo debido al fácil acceso a la compra e instalación de APs debido a ser de uso cotidiano comparado con otras tecnologías alternativas.

Más allá de la posibilidad de implementar una guía para los usuarios, se podrá

Capítulo 1. Introducción

recolectar información relevante para el museo tales como lugares u obras que atraen más visitantes o los días y horas con mayor frecuencia de visita. Esta información podrá ser de gran utilidad para la gestión de un espacio cultural como el MNAV.

Como tercer partícipe de este proyecto, se incorporó Antel como fuente de financiamiento y soporte logístico. Esta financiación incluye los APs, la instalación de los mismos en el museo y una beca para los integrantes del grupo de proyecto. Esta aproximación al problema resulta en un claro ejemplo de extensión universitaria, donde se trabajó en un marco en común entre Antel, el MNAV y la UdelaR.

1.1.2. Tecnologías de localización indoor

Las tecnologías de localización tienen un uso extendido a través de varios servicios disponibles para dispositivos móviles. Estos servicios suelen estar basados en GNSS (Global Navigation Satellite System), que constituyen sistemas de posicionamiento global basado en tecnología satelital. Sin embargo, esta tecnología cuenta con limitaciones para trabajar en ambientes cerrados, donde se tiene una precisión pobre que no permite una navegación correcta. El GNSS trabaja mejor si el dispositivo que lo utiliza tiene línea vista hacia el cielo. Esto se debe particularmente a que trabaja en frecuencias de 1575.42 MHz, teniendo problemas para propagarse a través de paredes y techos de un edificio.

Se han desarrollado alternativas a esta tecnología que permitan una correcta localización indoor (ambientes interiores), evitando las limitaciones. Una veta de desarrollo para estos sistemas se basa en el uso de señales de Wi-Fi para llevar a cabo la localización de dispositivos móviles tales como teléfonos inteligentes o tablets. Es en estas tecnologías que se basa el proyecto.

Las tecnologías de localización indoor tienen una gran cantidad de casos de uso. Por ejemplo, puede servir haciendo de sistemas de navegación para instalaciones como aeropuertos, museos y centros comerciales. Estos sistemas de navegación pueden verse complementados con aplicaciones específicas, tal como el caso del proyecto donde la información de posición del usuario forma parte importante de decidir el acceso a contenido del usuario. Todas estas entidades no sólo pueden tener un interés en proporcionar un servicio al usuario, sino en recolectar la información de los recorridos de los mismos. En el caso de los centros comerciales puede proporcionar una herramienta muy valiosa para analizar los distintos perfiles de consumidores a partir de las estadísticas de concurrencia de estos usuarios en distintas horas y días de la semana. Otro uso muy importante de esta información consiste en optimizar el flujo de personas tal como en un aeropuerto u otras instalaciones.

Otro tipo de casos de uso de localización indoor refieren al ámbito industrial, donde la trazabilidad de los distintos bienes, maquinaria e individuos puede ser clave para la optimización de los procesos. Este tipo de sistemas permite tener un constante registro de la ubicación de los distintos recursos, garantizando el rápido acceso a los mismos. Por otro lado, tener un sistema global que permita ubicar a todos los recursos en el tiempo establece un punto de partida para el diseño de

protocolos o procedimientos que maximicen la productividad.

1.2. Objetivos

Se busca implementar un sistema de localización indoor basado en Wi-Fi para el MNAV. Este permitirá mediante una aplicación de Android localizar al dispositivo dentro del museo.

En función de la posición del dispositivo, deberá desplegar información relativa a las obras dispuestas en la proximidad del usuario. Esta información se hará disponible en forma de imágenes, audio y texto, enriqueciendo la experiencia del visitante para la muestra Aquí soñó Blanes Viale.

La aplicación debe contar con una funcionalidad específica que permita el uso para personas con discapacidad visual, garantizándoles una mayor independencia y accesibilidad en el recorrido. Más allá de la implementación del sistema, se busca evaluarlo en torno a diversos escenarios de despliegue o funcionamiento.

1.3. Alcance

En la planificación inicial del proyecto, se definen los límites bajo los cuales se comprende el alcance del proyecto.

El área de localización indoor se encuentra en pleno desarrollo y tiene una gran variedad de aplicaciones. Se busca evaluar una implementación para posicionamiento indoor basado en señales Wi-Fi. En particular, el interés está centrado en marcos abiertos de uso libre.

La evaluación se centrará en la precisión que se obtenga (en el contexto de una muestra real en el MNAV) y su relación con la cantidad de los APs desplegados. Adicionalmente, la evaluación contará con un estudio de resolución espacial.

Por otro lado, se recolectará la información del uso de la aplicación entre los usuarios, generando estadísticas relevantes tales como zonas más visitadas y horarios de mayor actividad.

1.3.1. Accesibilidad

El aspecto quizás más novedoso con respecto a la aplicación de la localización indoor en el proyecto, consiste en el uso de esta tecnología para garantizar una mayor accesibilidad a las instalaciones del museo. Desde el punto de vista de la aplicación contar con la posición del usuario en el museo como una variable permite obviar interacciones mayores con la pantalla del dispositivo. Por otro lado, comparando al sistema anterior de Museo Amigo u otras alternativas comunes para garantizar mayor accesibilidad, un buen diseño de la aplicación evitaría la necesidad de apoyo en forma de otras personas o acompañantes, otorgando al usuario una mayor independencia.

Capítulo 1. Introducción

Dentro de las funcionalidades de la aplicación, será imprescindible contar con una alternativa al Museo Amigo en cuanto a accesibilidad para usuarios con discapacidad visual.

Capítulo 2

Estado del arte

2.1. Introducción

Previo a comenzar el diseño y la implementación del sistema, se realizó un estudio general de las tecnologías existentes para la localización indoor. Se estudiaron con un mayor énfasis aquellos sistemas que se consideraron factibles o pertinentes al proyecto, profundizando en la ingeniería detrás de los mismos.

Los sistemas de localización con mayor despliegue y uso a nivel global son basados en satélites: los denominados GNSS. Los más utilizados actualmente son el Sistema de Posicionamiento Global (GPS) de los Estados Unidos, el sistema Galileo de la Unión Europea y el Sistema Orbital Mundial de Navegación por Satélite (GLONASS) de la Federación Rusa. Estos sin embargo presentan inconvenientes al momento de localizar en ambientes cerrados, especialmente si se espera una precisión razonable para navegación en un edificio, museo o aeropuerto. Es entonces para estos casos de ambientes cerrados que se abre una gama de tecnologías diversas que apuntan a lograr un sistema de localización preciso y confiable. Como primer acercamiento se busca separar estos sistemas por la tecnología de fondo, haciendo énfasis en sus características y observando aquellos casos dónde más de una tecnología puede ser integrada.

2.2. Alternativas a WLAN para localización indoor

El posicionamiento indoor puede ser llevado a cabo basándose en diversos tipos de tecnologías. Más allá de que el proyecto se centró en el uso de señales de WLAN (Wireless Local Area Network), se estudiaron alternativas así como posibles complementos para el posicionamiento. Estas tecnologías presentan características muy disímiles, lo cual provee un gran rango de posibilidades para la combinación de las mismas.

Capítulo 2. Estado del arte

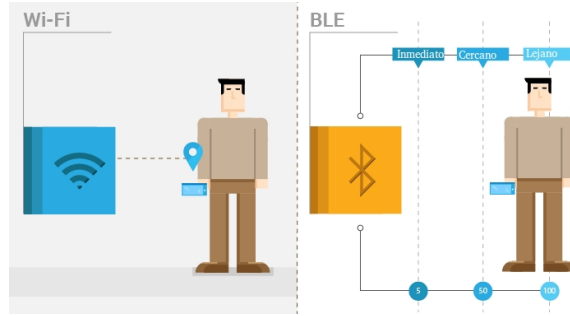


Figura 2.1: Mientras el posicionamiento con Wi-Fi suele buscar obtener la posición exacta, en BLE se busca obtener la proximidad a cada beacon. Figura obtenida de [10].

2.2.1. BLE

BLE (Bluetooth Low Energy), es una solución de Bluetooth pensada para un consumo mínimo de energía [22]. El desarrollo de esta tecnología ha permitido la introducción de los denominados beacons, dispositivos pequeños de bajo costo y consumo que generan señales de radio BLE. Como se verá cuando se discutan los sistemas basados en WLAN, el principio de funcionamiento de la localización con BLE es muy similar al de WLAN. Utilizando las potencias recibidas de los distintos beacons el móvil o dispositivo busca encontrar su ubicación. Sin embargo, la solución de localización con BLE puede ser basada en huellas de potencia o en proximidad.

Beacons BLE Los beacons de BLE presentan ventajas y nuevas posibilidades sobre las alternativas como los APs de WLAN. Al ser BLE de muy bajo consumo, estos beacons operan con baterías, por lo cual pueden ser instalados físicamente con menos restricciones que los APs. BLE está disponible en dos versiones [22], en la clase 1 tiene un rango de transmisión de hasta 100 metros en espacio abierto mientras que en clase 2 es de 10 metros. Para la localización se suelen utilizar con el rango de 10 metros ya que esto reduce el consumo del beacons así como el efecto de propagación multicamino presente por ejemplo en WLAN. Con este alcance menor al de WLAN, se requiere una mayor densidad para poder cubrir la misma área. A pesar de esto el costo de los mismos es de unas decenas de dolares haciéndolos muy accesibles.

Localización por proximidad La localización por BLE suele basarse en calcular proximidad. En este método se aprovecha el menor rango (que reduce el efecto de propagación multicamino) y la alta tasa de escaneo que permite promediar y eliminar puntos anómalos. Para localizar en modo proximidad, se mide el *RSS* (Received signal strength) de cada beacon y se lo compara con ciertos rangos. En distancias cortas (especialmente en menores a 1m), se puede modelar la caída del *RSS* con la distancia de forma precisa con la siguiente función [4]:

$$RSS(d) = RSS(d_0) + 10 n \log(d/d_0) + X_{\sigma}, \quad (2.1)$$

2.2. Alternativas a WLAN para localización indoor

donde d_0 es un punto de referencia, n es el exponente de atenuación (se suele considerar como 2 debido a que se puede considerarse como propagación en vacío) y X_σ es el ruido Gaussiano de varianza σ^2 con media nula.

En resumen, las tecnologías de BLE permiten una mayor exactitud que las de WLAN cuando se logra una densidad alta de beacons. Las diferencias en forma de funcionamiento y rendimiento con respecto a WLAN vuelven a la alternativa de BLE un posible complemento. En particular, el fácil despliegue de los beacons y el mejor rendimiento en distancias cortas permiten, por ejemplo, reforzar el sistema de localización en zonas de baja exactitud o difícil acceso. A pesar de esto se decidió no contar con esta tecnología dado que las compras de los mismos se deben hacer en el exterior agregándole entonces al proyecto un tiempo de espera para poder efectivamente probar la tecnología. Al usar tecnología de Wi-Fi rápidamente se pudo tener un sistema de pruebas funcional sin la necesidad de tener que incurrir en gastos. Es por esto que la tecnología Wi-Fi es mas conveniente, se pueden obtener sus beneficios al instante ya que hoy en día cualquier hogar cuenta con uno de ellos al menos.

2.2.2. Campo magnético

Otra tecnología utilizada para la localización indoor, se basa en estudiar la distribución del campo magnético dentro del edificio. Variaciones sobre el campo magnético terrestre dentro del ambiente son causadas principalmente por la estructura metálica del edificio [4]. Tal como en el caso de WLAN, se debe primero realizar un mapa en una etapa de entrenamiento, para luego a partir de las medidas obtenidas de valores de campo magnético inferir la posición del dispositivo. Esta localización es posible debido a que los smartphones suelen contar con un magnetómetro, el cual hace a su vez de elemento para la construcción del mapa de medidas y para la posterior medición para la localización.

Las características principales de este método de localización están dadas por la naturaleza del campo magnético en estos ambientes. Por ejemplo, el campo magnético es relativamente estable y se encuentra en todos lados. Estos aspectos significan en primer lugar un menor deterioro del mapa de huellas en el tiempo, ya que la señal se ve cambiada principalmente por cambios en la estructura del ambiente. Adicionalmente, la presencia del mismo naturalmente elimina la necesidad de instalación de infraestructura. Por otro lado, la medida del campo por ubicación tiene como máximo dos parámetros (direccionalidad y módulo), el gradiente del campo puede ser muy alto y las interferencias magnéticas intermitentes pueden ser muy significativas. Los pocos parámetros implican menores datos de comparación y aprendizaje. El gradiente elevado permite una localización de alta exactitud, sin embargo conlleva una necesidad de un mapa muy denso de huellas lo que aumenta en gran medida la dificultad de la etapa de mapeo siendo esta etapa una de las mas costosas en cuanto a horas hombre (como se pudo comprobar y se discutirá más adelante). Por lo tanto es por esto que esta alternativa no se utilizó. Una alternativa para la utilización del campo magnético es como complemento a otro método de localización como WLAN o BLE.

2.2.3. Acelerómetro y giroscopio

Una de las formas de complementar sistemas de localización es el uso de los sensores inerciales internos al móvil. A partir del acelerómetro y el giroscopio uno en principio podría obtener un diferencial de movimiento si se parte de una posición inicial ubicada con un método alternativo. Estos sensores no suelen ser muy precisos por lo cual se han diseñado sistemas en los cuales se evite una rápida acumulación de error. El método tiene que comprender distintas formas de transportes (ascensores, caminar, etc) y distintas personas con distintas cualidades físicas. Por lo cual estamos ante un método que no es robusto en cuanto a los cambios del sistema. Este también sin embargo pudo haber sido un elemento complementario a el uso de la tecnología WLAN, pero que no fue utilizado dado las restricciones temporales del proyecto.

Para asistir en este proceso, una técnica utilizada consiste en primero una clasificación del tipo de movimiento. De esta forma se pueden reconocer cambios de piso por ejemplo. Luego, para algunos casos tales como caminar, se busca estimar el patrón con el cual se camina (dirección y longitud de los pasos) y contar los mismos a través del análisis de los sensores. A partir de estos conceptos se puede ir agregando un delta de distancia en determinada dirección a partir de una localización inicial. Las técnicas para analizar y detectar estos patrones son diversas y se puede encontrar más en [4].

2.2.4. RFID

RFID (Radio Frequency Identification), utiliza ondas de radio para transmitir una identidad. Es común la utilización para identificación de objetos y en particular el uso con dispositivos pasivos de RFID. Estos sistemas son muy utilizados a distancias pequeñas. En el funcionamiento pasivo de RFID existe un tag pasivo que solo responde ante una señal del RFID reader.

Para la localización indoor, hay dos metodologías principales. En una primera, se busca localizar al tag, mientras que en la segunda se buscan localizar RFID readers portátiles. En principio tanto las tags como los lectores no son objetos o equipos comunes como los APs de Wi-Fi y los smartphones, por lo cual la localización parte de que se precisa una nueva infraestructura. Más allá de sus interesantes aplicaciones quizás para un ámbito industrial, esta tecnología no se ajusta a los objetivos o posibilidades del proyecto.

2.3. Localización basada en WLAN

Los sistemas basados en WLAN buscan obtener información sobre la localización a partir de las señales generadas por puntos de acceso WLAN (APs). Este método es muy conveniente, ya que es muy común contar con una infraestructura de Wi-Fi y los dispositivos tales como smartphones y tablets soportan esta tecnología.

2.3.1. Received Signal Strength Indication: RSSI

La principal forma de inferir la ubicación del dispositivo es a partir de la potencia recibida de cada AP. El RSSI (Received Signal Strength Indication) es una medida de la potencia relativa de la señal recibida luego de las posibles pérdidas de antena y línea [15]. El indicador trabaja con unidades arbitrarias y no está definida una relación entre la RSSI y potencia en mW (mili-Watts) o dBm (Decibeles referidos a un mW). La diferencia principal entre *RSS* y RSSI es que la primera es el valor real de potencia recibido, mientras que RSSI es arbitrario en cuanto a unidad y escala.

En la práctica, a partir de la medida de RSSI se suele obtener una medida relativa de la potencia recibida lo cual permite dar una idea de la proximidad o ubicación de los APs. Sin embargo, se debe tener en cuenta que la medida no es el *RSS* y depende, por ejemplo, de la antena y el chip del dispositivo.

2.3.2. Beacon frames

Los APs transmiten en forma periódica, típicamente cada 102.4ms, “beacon frames” que contienen la dirección MAC única correspondiente al AP [11]. A partir de la dirección MAC se puede entonces identificar al AP que genera la señal, con lo cual se pueden obtener las potencias recibidas de cada AP. Como aproximación intuitiva, tener los valores de potencia recibida puede dar una idea de la cercanía con respecto al AP en cuestión.

Los “beacon frames” son tramas que están definidas acorde a las especificaciones de la IEEE para 802.11 [11]. Estas tramas están compuestas por un encabezado de MAC (de 802.11), el cuerpo del mensaje y un código de corrección de errores FCS (Frame check sequence). El cuerpo del mensaje incluye los siguientes campos [11]:

- Timestamp: Permite sincronización.
- Beacon interval: Tiempo entre beacon frames.
- SSID: Permite identificar la red más allá del AP en particular.

Existen en total al menos otros 65 campos que pueden estar presentes en el beacon frame dependiendo del uso, de menor relevancia en la localización indoor.

2.3.3. Escaneo activo y pasivo

Esta primera alternativa a la localización por WLAN utiliza una forma de escaneo activo. Este término refiere a que el dispositivo a ser localizado reporta o utiliza las potencias de las señales de los APs circundantes para obtener su propia ubicación.

Por contraposición, la localización por escaneo pasivo busca ubicar al dispositivo mediante distintas lecturas de potencia de la señal emitida por el mismo. Es decir, en este tipo de escaneo, los APs pasan ser los que miden la potencia y el

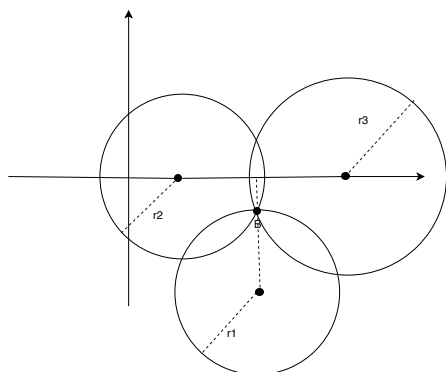


Figura 2.2: Trilateración geométrica.

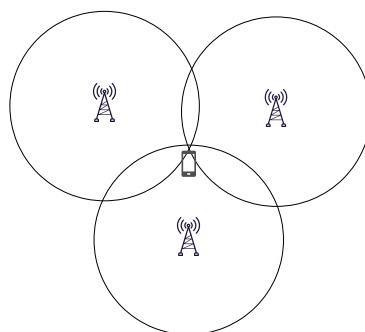


Figura 2.3: Trilateración a partir de medidas de *RSS*.

dispositivo pasa a ser el generador de la señal a medir. Este segundo método requiere incorporar un sistema distinto. Como en esta metodología los APs tienen la función de medir RSSI y reportarlo a algún servidor central, tienen que contar con una inteligencia adicional. La localización pasiva tiene como ventaja de que retira la necesidad de que el móvil lleve a cabo la localización, por lo cual no necesita de una aplicación.

2.3.4. Trilateración

Una primera forma de localizar el dispositivo a partir de los valores de *RSS* es utilizando trilateración [4]. Para ello, primero se debe inferir una medida de distancia del dispositivo a los APs a partir del RSS. Luego de obtenida la distancia aproximada a cada AP, se puede utilizar la trilateración para obtener la posición relativa del dispositivo con respecto a los APs como en la figura 2.3 y 2.2.

Sin embargo, este prueba ser un sistema muy poco preciso y de gran inestabilidad. Esto se debe a que las medidas de *RSS* tienen una variabilidad espacial considerable, por lo cual al buscar obtener distancias directamente a partir de ellas, se obtiene una medida con error. Este fenómeno es incluso peor en los ambientes cerrados debido a que las señales de Wi-Fi pueden ser atenuadas y pueden rebotar en las paredes y objetos. Por estas razones, este método es descartado.

2.3.5. Sistema de huellas digitales

En contraposición al método de trilateración, otra alternativa es proponer un modelo que capture las características de la propagación sin tener en cuenta las dificultades de un modelo complejo de propagación. Es aquí donde se introduce el método de huellas digitales. Este es el método más utilizado para la localización indoor en tecnologías WLAN. Un sistema de localización basado en huellas digitales consta de dos etapas. Una primera etapa de aprendizaje offline, se miden los valores de *RSS* para distintos puntos del local recibidos de varios APs. De esta forma se genera un mapa de *RSS* o 'huellas' asociadas a posiciones en el local. Podemos ver en la figura 2.5 la representación gráfica de como esta caracterizado

2.3. Localización basada en WLAN

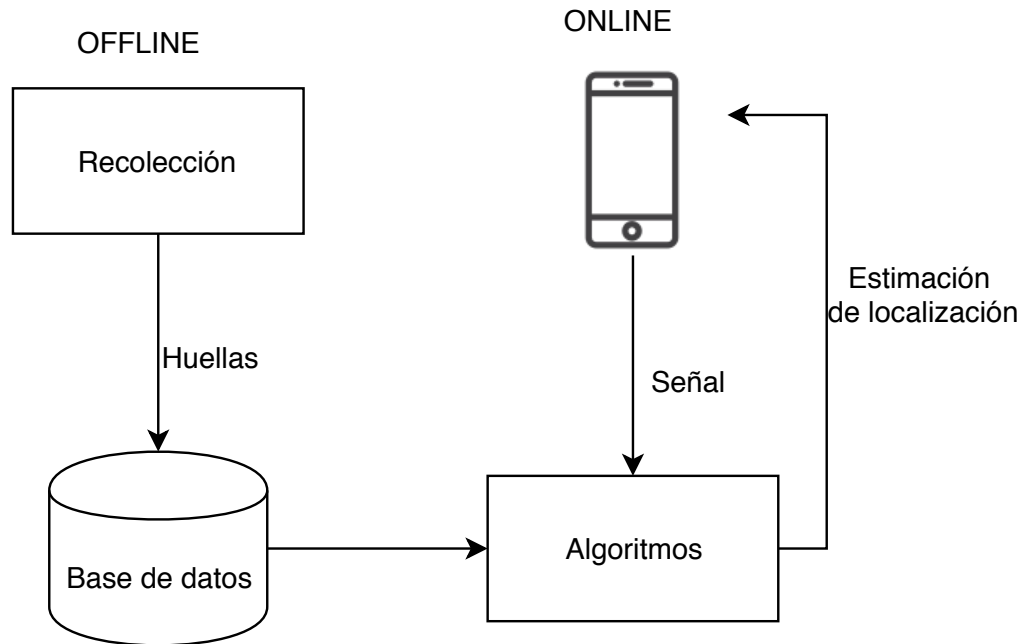


Figura 2.4: Flujo básico de Sistema utilizando huellas.

un punto por su huella. Y es partir de estas huellas tomadas inicialmente que se entrena un modelo que luego se utilizara para estimar la posición de las futuras huellas. En la segunda etapa, un móvil mide valores de *RSS* de los APs comparándolos con el modelo de huellas obtenido en la primera etapa. Esta etapa es la etapa de clasificación de las huellas donde a partir del modelo realizado en la primera etapa se intenta predecir la localización a partir de las medidas de nuevas huellas. Podemos ver una descripción gráfica de el sistema en la figura 2.4

Mapeo RSS

El mapeo de *RSS* es el método por el cual se obtienen las 'huellas' para la posterior localización de los dispositivos. En este mapeo lo que se realiza es la división del plano del local en zonas. Luego, se miden las *RSS* obtenidas de los distintos APs, para cada una de estas zonas. Cada huella es entonces un vector que en cada uno de sus elementos posee un valor de la señal de *RSS* de un AP. Es decir, a cada zona se le asigna un vector con las *RSS* obtenidas de los distintos APs. Estos vectores de *RSS* son los que servirán de referencia en la etapa de localización, donde se busca ubicar en qué zona se encuentra el dispositivo. Una de las mayores desventajas del método radica en esta etapa de mapeo o generación de las 'huellas'. La generación del mapa de *RSS* es laborioso y consume tiempo. Adicionalmente, este mapa se deteriora con el tiempo ya que cambios en el entorno o en la disposición de los APs suelen hacer que los valores de *RSS* para cada posición cambien con el tiempo. Esto implica entonces un mantenimiento adicional del mapa para no perder precisión en el sistema utilizado.

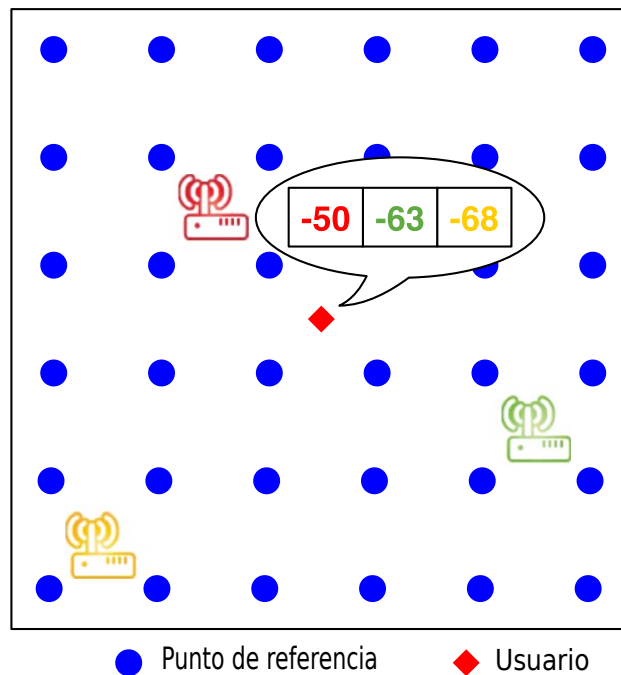


Figura 2.5: Mapeo del sitio con valores de RSS representados en dBm [15]

2.4. Clasificación y aprendizaje utilizando huellas

Una vez realizado el relevamiento del mapa de RSS, existen diversas formas de utilizar esos datos. Lo que se busca en términos básicos es comparar el vector de RSS obtenido por un dispositivo con los valores del mapa y a partir de la información del mapa decidir en que posición es más probable que se encuentre. Hay varias posibilidades para lograr este objetivo con distintas características, pero básicamente este problema se puede traducir a un problema de clasificación de huellas.

2.4.1. Algoritmos

Como estamos frente a un problema de clasificación podemos entonces plantear observar los distintos algoritmos que existen para resolver esta problemática. En un problema de clasificación se busca identificar a que conjunto de categorías pertenece una nueva muestra, basándose en un conjunto de entrenamiento que contiene muestras ya asignadas. Es decir, en este caso se busca identificar a que posición pertenece una nueva muestra (vector de RSS obtenido por dispositivo), a partir de un conjunto de entrenamiento (mapa de RSS) que contiene muestras (vectores de RSS) ya asignados (con posiciones específicas).

La mayoría de los algoritmos busca minimizar una distancia entre el nuevo vector de RSS con aquellos pertenecientes al mapa. De esta forma, buscar obtener una noción de cual es el punto más cercano en esta definición de distancia.

Para ser mas precisos se tienen distintas clases que representan las zonas, y

2.4. Clasificación y aprendizaje utilizando huellas

una vez que se sabe la distancia más cercana el vector se etiqueta como de esa clase. Los algoritmos intentan aproximar una función capaz de predecir el valor correspondiente a cualquier vector de RSS de entrada válida después de haber visto el mapa de RSS , los datos de entrenamiento. Esta función a la salida devuelve una probabilidad de que el elemento de entrada pertenezca a las distintas clases.

Algunos de los algoritmos clásicos que se utilizan para este tipo de problemas en reconocimiento de patrones son: KNN (K-Nearest Neighbors), WKNN (Weighted KNN), SVM (Support Vector Machines), Decision Trees, Random Forests y redes neuronales.

KNN y WKNN

Algunos de los métodos más usados provienen de la familia de algoritmos de “nearest neighbour” (vecino más cercano). Sirve esencialmente para clasificar valores buscando los puntos de datos “más similares”. La fase de entrenamiento del algoritmo consiste únicamente en almacenar los vectores de característica y las etiquetas de clase de las muestras de entrenamiento. Los datos de entrenamiento son vectores en un espacio de características, cada uno con una etiqueta de clase. En la fase de clasificación, k es una constante definida por el usuario y un vector no identificado (punto de prueba) se clasifica asignando la etiqueta más frecuente entre las muestras de entrenamiento k más cercanas a ese punto de consulta. Es por eso que la k significa la cantidad de “puntos vecinos” que tenemos en cuenta en las cercanías para clasificar las distintas clases. El valor de k es un parámetro de diseño a elegir. Para un mapa denso, k debe ser bajo, alrededor de 3 o 4 es el óptimo en la mayoría de estos casos [4]. Luego, se obtiene la posición final a partir de estos “puntos vecinos”. En KNN, estos puntos tienen todos la misma importancia para computar la posición, independientemente de cuanto era la distancia al vector medido. En el caso de WKNN (Weighted KNN), se le asignan distintas importancias a los vecinos en forma de pesos. Esto permite, por ejemplo, darle más importancia a los puntos más cercanos al vector.

Estos métodos presentan muchas ventajas más allá de la simplicidad. En las implementaciones existentes han demostrado una buena exactitud y una robustez ante problemas como la degradación del mapa de RSS [4].

SVM (Support Vector Machines) [3]

Las Support Vector Machines (Máquinas de vectores de soporte) son modelos ampliamente utilizados para resolver problemas de clasificación. SVM es un clasificador lineal dado que a partir de tener los datos como puntos del espacio el algoritmo encuentra una superficie o hiper-plano que separe los datos de forma óptima para poder tomar una decisión de clasificación. Esto se ilustra mejor en la figura 2.6. El hiperplano generado, separa el espacio en regiones, de forma tal que los puntos son clasificados acorde a la región a la cual pertenecen. Los nuevos datos se mapean a este espacio y según la región en la que caigan se los clasificará.

Capítulo 2. Estado del arte

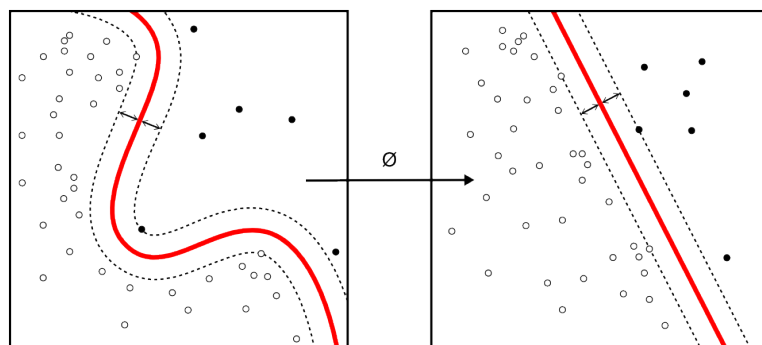


Figura 2.6: Las máquinas de vectores de soporte mapean los vectores no linealmente separables en un espacio donde esto si sea posible [1].

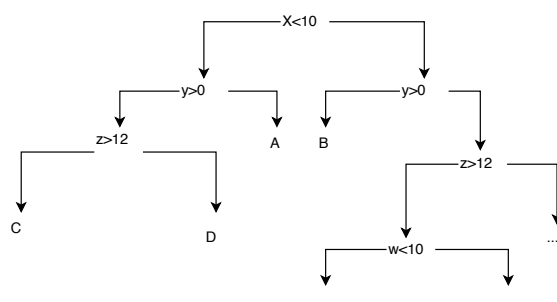


Figura 2.7: Árbol de decisión. En cada uno de los nodos se realiza alguna prueba y los resultados de las mismas conducen a los nodos hoja que corresponden a la clasificación final.

Arboles de decisión [16]

Árbol de decisión es un algoritmo que utiliza un modelo de decisiones similar a un árbol en el cual los nodos internos son pruebas sobre las entradas y los nodos “hoja” (las puntas inferiores del árbol) son los posibles resultados de la clasificación. Esto se ilustra en la figura 2.7, donde (X,Y,Z,W) son parámetros de los datos de entrada. En cada nodo partiendo desde la raíz se verifica una condición con respecto a la entrada, conduciendo eventualmente a una decisión, como en este caso son A,B,C y D. Como variante más general, la salida de los nodos internos no tiene porque ser binaria y pueden abrirse más de dos ramas. En el aprendizaje por árboles de decisión, se busca determinar un árbol a partir de algoritmos donde se determina la estructura en cada nodo. Existen varios procedimientos para poder decidir el largo de los arboles de decisión. Una forma fácil de hacerlo es establecer un número mínimo de elementos de entrenamiento para usar en cada hoja. En general el procedimiento que se utiliza para determinar las pruebas en cada nivel es usar la idea de costo de hacer una división. Se busca entonces en reducir el costo después de que un conjunto de datos se divide en un atributo. Entonces el objetivo es minimizar este costo.

2.4. Clasificación y aprendizaje utilizando huellas

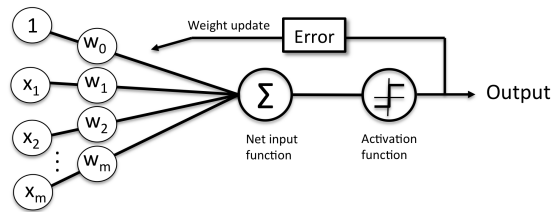


Figura 2.8: Perceptrón [19]. A partir de la diferencia entre la salida esperada y la salida del algoritmo se realiza una actualización de los pesos para la combinación lineal.

Random forests [7]

Random Forests (o en castellano, Bosques Aleatorios) es una técnica de aprendizaje automático derivada de los árboles de decisión. En este método, se busca crear muchos árboles de decisión distintos. Estos distintos árboles se ensamblan (usualmente con el método conocido como “bagging”) de forma tal de obtener un mejor resultado. La idea general del método de “bagging” es que la combinación de métodos de aprendizaje obtiene mejores resultados que cualesquiera de ellos individualmente. Los Bosques aleatorios agregan un factor de aleatoriedad al construir los árboles de decisión. El árbol de decisión suele buscar la característica óptima a probar de los datos de entrada para la definición de cada nodo. En los Bosques aleatorios, se busca esta característica solamente entre un subconjunto aleatorio de todas las características. Incluso se puede agregar otro factor de aleatoriedad en la forma de límites aleatorios en lugar de óptimos para la elaboración de las condiciones de los nodos. Este aspecto agrega una mayor variedad en la creación de los árboles que generalmente resulta en un mejor modelo.

MLP (Multi-layer perceptron)

Un perceptrón es un algoritmo de aprendizaje automático para clasificadores binarios inspirado en el funcionamiento de una neurona [19]. Las neuronas actúan como compuertas lógicas simples con salida binaria, donde las entradas corresponden a las señales químicas y eléctricas recibidas. En el caso de que la señal acumulada supere cierto rango se genera una señal de salida que va a ser transmitida. De la misma manera, el perceptrón tiene una estructura simple, donde a partir de una combinación lineal de n entradas y una función de activación, se obtiene una salida binaria. La clave del algoritmo es hallar los pesos óptimos de la combinación lineal para la mejor clasificación. Este proceso se realiza entrenando al algoritmo con un set de entrenamiento, actualizando los pesos de la combinación lineal a partir de la diferencia entre la salida del perceptrón y de la muestra. Este proceso se ilustra en la figura 2.8. Dada la simplicidad del perceptrón, este solo es capaz de resolver problemas linealmente separables.

Un Multi-layer perceptron (en castellano perceptrón Multi-capas), parte del perceptrón para crear una red neuronal más compleja. El propósito de esto es poder resolver problemas de clasificación no linealmente separables [19]. En esta

Capítulo 2. Estado del arte

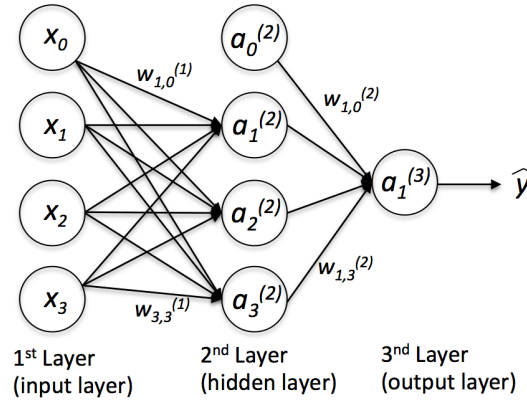


Figura 2.9: Perceptrón Multi-capas [19]. La añadición de la capa intermedia le da una mayor capacidad y potencial para desarrollar problemas de mayor complejidad.

red existe al menos una capa intermedia entre la entrada y la salida como capa oculta. Esta capa consiste conceptualmente de varios “perceptrones simples”, todos alimentados por las entradas al sistema. Finalmente la combinación de las salidas de esta capa pasan por una función de activación y dan el resultado final. El entrenamiento de este algoritmo consiste entonces en optimizar todos los pesos involucrados de las combinaciones para llegar a la mejor clasificación. La estructura de la red neuronal se observa en la figura 2.9. El entrenamiento de la red se suele realizar mediante algoritmos como el de propagación hacia atrás.

AdaBoost [26]

AdaBoost es un meta-algoritmo de aprendizaje automático basado en la idea de que contar con un grupo de algoritmos para tomar decisiones es mejor que tener uno solo. Un meta-algoritmo es un algoritmo que utiliza uno o varios sub-algoritmos para aprender. La experiencia se obtiene mediante la explotación de los datos extraídos en un episodio de aprendizaje anterior en un único conjunto de datos, o de diferentes dominios. En el caso de AdaBoost al grupo de algoritmos se le conoce como ensamble y este representa un clasificador fuerte, es decir, un clasificador con una precisión muy buena. Por su parte, a los algoritmos que conforman el ensamble se les denomina clasificadores débiles, es decir, clasificadores con una precisión menor.

El nombre de este meta-algoritmo proviene de la abreviación de adaptive boosting y hace referencia a una estrategia de ponderar de manera diferente a cada uno de los ejemplos, durante el entrenamiento. Aquellos ejemplos que fueron clasificados incorrectamente en una iteración, reciben una ponderación más alta en la siguiente iteración. Por el contrario, los ejemplos que fueron clasificados correctamente en esa iteración, recibirán menos importancia en la siguiente. Esta estrategia permite que AdaBoost se enfoque, iteración tras iteración, en aquellos ejemplos

2.4. Clasificación y aprendizaje utilizando huellas

que aún no han sido clasificados correctamente por el ensamble de clasificadores débiles.

QDA (Quadratic discriminant analysis)

QDA es una generalización de Linear discriminant analysis (análisis por discriminante lineal en castellano). Este es un método de clasificación donde se busca caracterizar o separar grupos mediante una combinación lineal de las entradas [18]. Para realizar esto, se utiliza el Teorema de Bayes. Sea $RSSv$ el vector de valores RSS medido y x la zona donde se encuentra el móvil, el uso de Bayes resulta en la siguiente ecuación (2.2):

$$p(x|RSSv) = \frac{p(RSSv|x)p(x)}{p(RSSv)}. \quad (2.2)$$

La probabilidad condicional $p(x|RSSv)$ en la ecuación (2.2) es la probabilidad de que el móvil se encuentra en la zona x habiendo medido un vector de potencias $RSSv$, mientras que $p(RSSv|x)$ es la probabilidad de medir un vector de potencias $RSSv$ estando el móvil en la zona x .

Sea d la cantidad de características (features) y k una clase dada en la clasificación (en este caso una zona). En LDA se modela a la probabilidad $p(RSSv|x)$ como una distribución Gaussiana multi-variable con densidad dada por:

$$p(RSSv|x = k) = \frac{1}{(2\pi)^{d/2}|\sigma_k|^{1/2}} \exp -\frac{1}{2}(RSSv - \mu_k)^t \sigma_k^{-1} (X - \mu_k), \quad (2.3)$$

donde σ_k y μ_k representan la matriz de covarianza y la media para la zona k . Se estima la probabilidad a priori $P(x = k)$, la media μ_k y las matrices de covarianza σ_k a partir de las muestras de entrenamiento. En el caso de LDA, se asume que las Gaussianas de cada zona comparten la misma matriz de covarianza. Esto resulta en superficies de decisión lineales, como se observa en la figura 2.10.

Luego para QDA (Quadratic discriminant analysis o análisis discriminante cuadrático) se levanta el supuesto sobre las matrices de covarianza. Esto resulta en superficies de decisión cuadráticas y por tanto más versátiles como podemos observar en la figura 2.10.

Clasificador bayesiano ingenuo

Los métodos bayesianos ingenuos son un conjunto de algoritmos que parten de aplicar el teorema de Bayes. Luego toman supuestos “ingenuos” de independencia condicional entre cada par de características de entrada, dada la clase (en localización, dada la zona) [18].

La suposición de la independencia condicional en el caso de localización se puede expresar como:

$$p(RSS_i|x, RSS_1, ..., RSS_{i-1}, RSS_{i+1}, ..., RSS_n) = p(RSS_i|x). \quad (2.4)$$

Capítulo 2. Estado del arte

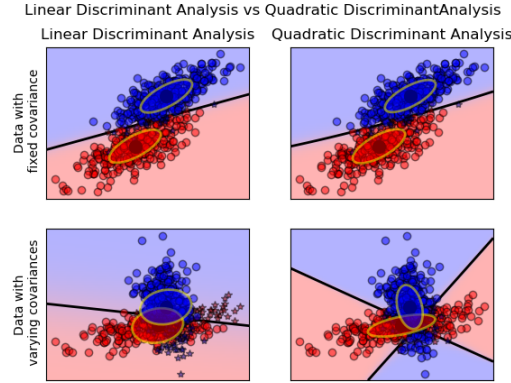


Figura 2.10: LDA y QDA utilizados en un mismo set de datos [18]. El levantar la suposición de tener las mismas matrices de covarianza genera que las fronteras de decisión sean cuadráticas y por tanto más flexibles.

Esto es suponer que la probabilidad de recibir una potencia RSS_i de un AP dada la posición no cambia dadas las otras RSS_i .

Utilizando esta suposición en la ecuación de Bayes (2.2), se tiene:

$$p(x|RSS_v) = \frac{p(x) \prod_{i=1}^n p(RSS_i|x)}{p(RSS_v)}. \quad (2.5)$$

Como la probabilidad $p(RSS_v)$ es constante dadas las potencias, se llega a la relación:

$$p(x|RSS_v) \propto p(x) \prod_{i=1}^n p(RSS_i|x), \quad (2.6)$$

por lo tanto el clasificador debe hallar la estimación de x tal que se maximice la expresión de la derecha.

$$\hat{x} = \underset{x}{\operatorname{argmax}} p(x) \prod_{i=1}^n p(RSS_i|x). \quad (2.7)$$

Por un lado $p(x)$ se puede estimar a partir de la proporción de huellas por zona en el set de entrenamiento (suponiendo que se emuló el comportamiento típico de un usuario real). Por otro lado, el estimador dependerá de como se asuma la distribución $p(RSS_i|x)$, por ejemplo, una distribución Gaussiana.

2.5. Combinación de algoritmos para mejorar exactitud

2.5.1. Motivación

Para el problema de clasificación planteado en localización indoor con WLAN, el algoritmo Random Forest, por ejemplo, logró una exactitud promedio de 75 %

2.5. Combinación de algoritmos para mejorar exactitud

según lo reportado en [25]. Es razonable y factible utilizar entonces tan sólo este algoritmo. Sin embargo, se pueden aprovechar los otros algoritmos, buscando explotar las ventajas y características de cada uno en una combinación resultados. Esto mejora el rendimiento predictivo mas allá del que podría obtenerse de cualquiera de los algoritmos de aprendizaje constituyentes por sí solo.

Esto se llama ensamble de algoritmos y permite que exista una estructura mucho más flexible a la hora de elegir los algoritmos. Los ensambles combinan múltiples hipótesis para formar una (en teoría) mejor hipótesis.

La evaluación de la predicción de un ensamble requiere típicamente más computación que la evaluación de la predicción de un solo modelo, por lo que los ensambles pueden ser considerados como una forma de compensar los pobres algoritmos de aprendizaje mediante la realización de una gran cantidad de cálculos adicionales. Los algoritmos rápidos como los árboles de decisión se utilizan comúnmente en los métodos de ensamblaje (por ejemplo, los bosques aleatorios), aunque los algoritmos más lentos también pueden beneficiarse de las técnicas de ensamblaje. Esto fue así para el caso del framework utilizado, pero que luego de la experiencia se decidieron por aquellos mas performantes y con menor tiempo de respuesta como veremos más adelante.

2.5.2. Índice de Youden

Para lograr la combinación de los algoritmos que mejorara el rendimiento del sistema, [25] basó esta combinación en una métrica llamada Índice de Youden [27]. El Índice de Youden se utiliza para obtener una medida del rendimiento de una prueba de diagnóstico dicotómica. Las pruebas de diagnóstico dicotómicas son aquellas donde existen dos resultados posibles, positivo o negativo.

El Índice de Youden se define como:

$$J = \text{especificidad} + \text{sensibilidad} - 1. \quad (2.8)$$

Para dar entonces una mejor idea del rendimiento, se definen otros valores como la especificidad y la sensibilidad.

Los conceptos de especificidad y sensibilidad vienen definidos de la siguiente manera:

$$\text{especificidad} = \frac{VP}{VP + FN}. \quad (2.9)$$

$$\text{sensibilidad} = \frac{VN}{VN + FP}. \quad (2.10)$$

Estos dos dan una idea del rendimiento de los algoritmos. Para darle un significado a las siglas que definen estos dos conceptos podemos ver que siempre se puede realizar un diagnóstico a partir de un conjunto de entradas cuyos resultados reales sean conocidos. Los resultados de comparar el diagnóstico y el valor real conducen a cuatro casos posibles, obtenidos a partir de la combinación entre los posibles diagnósticos y valores.

Capítulo 2. Estado del arte

- (VP) : Verdaderos positivos, son aquellos resultados que están clasificados correctamente como positivos.
- (VN) : Verdaderos negativos, son aquellos resultados que están clasificados correctamente como negativos.
- (FP) : Falsos positivos, es cuando se recibe un resultado positivo para una prueba, cuando se debería haber recibido un resultado negativo.
- (FN) : Falsos negativos, es cuando se recibe un resultado negativo, pero se debería tener un resultado positivo.

A partir de los resultados de la prueba hay varias formas de evaluar el rendimiento del predictor. Entre ellas la más conocida e intuitiva es la exactitud. Es decir, cuantos resultados de diagnóstico fueron correctos sobre el total.

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN}. \quad (2.11)$$

De aquí también se desprenden las definiciones para la especificidad y la sensibilidad. Entonces teniendo estas definiciones podemos decir que la sensibilidad mide la proporción de casos positivos que se identifican correctamente como tales. Mientras que la especificidad mide la proporción de negativos reales que se identifican correctamente como tales.

Incluyendo las definiciones de especificidad y sensibilidad en la ecuación se tiene:

$$J = \frac{VP}{VP + FN} + \frac{VN}{VN + FP} - 1. \quad (2.12)$$

El valor puede ser de -1 a 1 y resulta en una medida del rendimiento del diagnóstico, siendo 1 la prueba perfecta. Para evaluar el resultado del índice en el caso de uso, se utiliza un dataset de prueba sobre el cual evaluar los distintos algoritmos.

Para lograr entonces el ensamble, usamos entonces el Índice de Youden. A partir de este Índice se pueden ponderar los resultados de los algoritmos utilizados para la clasificación. Definimos $J_{\omega y}$ como el Índice de Youden determinado para cada algoritmo ω , en cada ubicación y , y la probabilidad $P_{\omega}(y)$ como la probabilidad de pertenecer a la ubicación y dada por el algoritmo ω . Entonces se puede obtener una probabilidad resultante del algoritmo ponderada de la forma:

$$Q_{\omega}(y) = J_{\omega y} P_{\omega}(y). \quad (2.13)$$

Luego de obtener esta medida para cada algoritmo, se combinan los algoritmos acorde a sus rendimientos con una simple sumatoria:

$$Q(y) = \sum_{\omega=1}^N Q_{\omega}(y). \quad (2.14)$$

Una vez que se obtiene este valor, a la salida de este estimador tenemos un conjunto ordenado de valores normalizados de $Q(y)$ donde el valor más alto es probablemente la mejor respuesta.

2.6. Consideración de la variable temporal

En particular, en la experiencia en cuestión [25], utilizando una combinación de los algoritmos, se logró una exactitud promedio de 87 %, lo cual presentó una mejora notoria con respecto al 75 % obtenido utilizando solo Random Forests. En los resultados finales del sistema que se puso a prueba se pudo llegar hasta el 96 % de exactitud utilizando este método, por lo cual el que el framework haya elegido usar este ensamble fue altamente ventajoso para que el éxito de la solución fuera posible.

2.6. Consideración de la variable temporal

El problema de localización planteado como un problema de clasificación a partir de las huellas *RSS* en principio no hace uso de la información sobre la ubicación pasada del móvil. Es decir, en principio el problema está planteado como una estimación estática de la posición. Sin embargo, existen diversas formas de hacer uso de la información pasada, incluyendo de esta forma la variable temporal en la estimación de la posición. De esta forma se lograría una estimación dinámica de la posición. Para incorporar esta información se utiliza generalmente un filtro Bayesiano o filtro de Kalman.

2.6.1. Filtro Bayesiano

La utilización del filtro Bayesiano busca incorporar la variable temporal a la estimación por máxima verosimilitud. La estimación de máxima verosimilitud viene definido por el enfoque probabilístico o bayesiano. En este enfoque la posición a estimar se asume una variable aleatoria y se busca que la probabilidad condicional, es decir aquella a posteriori, sea máxima a partir de la información de las huellas. A partir de esto se extiende esta definición que en este caso sería para la posición estática, agregando la información del pasado de la trayectoria. En este caso el modelo de movimiento de los dispositivos se define en la elección del prior. Esta era una de las elecciones que tenían más sentido para implementarse en la solución final, sin embargo al final con las limitantes de tiempo no se pudo implementar y se decidió por una solución mas hecha a medida (ver sección 3.6.3).

2.6.2. Filtro de Kalman

El filtro de Kalman es un caso particular del filtro Bayesiano. En este, se modela linealmente el movimiento del móvil, agregando un ruido blanco gaussiano de media nula. Existen a su vez derivaciones del filtro de Kalman como PKF explicada en [9] y FKF presentada en [8] para el uso en el problema de localización. PKF (Position Kalman Filter) se usa para suavizar las medidas y cambios bruscos de posición obtenidos en la solución estática. Este filtro utilizaría las soluciones estáticas como medidas originales para el filtro de Kalman convencional. Por otro lado FKF (Fingerprint Kalman Filter), depende menos de la exactitud de la solución estática. Este método se basa en la estimación recursiva de BLUE (Best

Capítulo 2. Estado del arte

Linear Unbiased Estimator), estimador lineal insesgado óptimo. En [8], se prueban los distintos filtros utilizando WKNN y se concluye que FKF obtiene una mejor exactitud y un menor consumo de recursos de procesamiento y memoria.

2.7. Elección y descripción de la arquitectura del sistema

Como se mencionó anteriormente se decidió trabajar con un framework que resolviera el problema de clasificación de huellas *RSS* de tal modo que se pudiera abordar el problema de forma más eficiente y rápida. Se decidió buscar frameworks de código abierto.

Se investigaron distintas alternativas y se pudo reducir las opciones a dos posibles. Una de ellas Framework for Internal Navigation and Discovery (FIND3) [25] y la otra AnyPlace [24].

Luego de comparar ambas, se observó que ambas cumplían con los requerimientos que se planteaban para el proyecto en cuanto a resolver el problema de posicionamiento en ambientes internos. Sin embargo, a la hora de tomar la decisión, se eligió trabajar con FIND3 dado que las tecnologías que maneja son acordes a las habilidades presentes en los integrantes del grupo. Esto es debido principalmente a que el lenguaje de programación en el que Anyplace está desarrollado es Scala, el cual ninguno de los integrantes manejaba. Por otro lado, FIND3 se encuentra programado en Python y en GO. Los integrantes del proyecto tienen experiencia con Python y no era necesario ahondar demasiado en aprender GO para entender el código en el framework.

Adicionalmente, se observó que la adaptabilidad de FIND3 era mayor, dado que la tecnología con la cual está diseñado permite que los distintos componentes se puedan modificar como se explicará mas adelante. Por estos dos motivos principales es que FIND3 se adecuaba mejor al proyecto, brindando mayor flexibilidad y mejor adaptabilidad a las necesidades del mismo.

Otro de los problemas que se tuvo que solucionar fue la elección de una aplicación en un dispositivo móvil para crear la interfaz del usuario con la cual va a navegar y recibir información dentro del museo. Para resolver este problema también se planteó buscar alternativas de código abierto para poder adaptar a la solución. Se buscaron opciones que fueran para la plataforma Android, buscando así que la aplicación tenga el mayor alcance de los usuarios finales.

Se investigaron cuales eran las interfaces que habían disponibles de código abierto. La primera solución encontrada es una aplicación de audioguía inclusiva del Museo Andy Warhol [6]. A través de la aplicación uno puede aprender sobre la vida y el arte del artista pop Andy Warhol a través de diferentes historias y una variedad de voces, como académicos, curadores y miembros de la familia de Warhol.

La aplicación está diseñada para incluir a los usuarios con distintas capacidades. Para los usuarios ciegos o con baja visión, la aplicación está optimizada para trabajar con entradas dinámicas. La aplicación también ofrece descripciones

2.7. Elección y descripción de la arquitectura del sistema

visuales detalladas de las obras de arte de Warhol.

Esta aplicación encajaba perfectamente con lo que se buscaba en cuanto a términos de interfaz. Sin embargo, usa los beacons de BLE y no está adaptada para hacer uso de la tecnología de Wi-Fi, siendo esto una limitante a la hora de ganar tiempo usando una interfaz ya diseñada. Por otro lado, como se mencionó anteriormente se estaba buscando una aplicación para la plataforma Android, y esta aplicación solo está presente para IOS. Por estos motivos se tuvo que buscar otra alternativa.

Luego de seguir investigando, se pudo encontrar una aplicación de guía turísticas del museo de arte Craft and Folk [2], que brinda la solución para Android, y que cuenta con una audioguía. El único cambio que debía hacerse era la incorporación del framework ya que esta contaba con un flujo muy parecido al que se quería llegar. Es por esto que finalmente se decidió adaptar esta última a las necesidades del proyecto. Esta decisión ayudó a que finalmente el despliegue de la aplicación fuera verdaderamente un proceso mucho más ágil y fácil de implementar, por lo cual decidirse por iterar por encima de un framework fue una buena decisión a la hora de la planificación.

2.7.1. Framework for Internal Navigation and Discovery: FIND3

FIND3 es un framework que permite usar un smartphone Android u ordenador con Wi-Fi para determinar su posición dentro de una casa, oficina u otro ambiente cerrado. Este framework utiliza las bases de localización indoor mediante WLAN y BLE vistas previamente. Mediante este framework, se puede en principio recolectar huellas de *RSS* y luego ubicar la posición del smartphone o dispositivo a partir de las mismas.

A diferencia del GPS, la aplicación utiliza operaciones preexistentes del teléfono, por lo que no tiene un costo significativo para la batería. Se puede utilizar fácilmente este sistema en lugar de los sensores de movimiento del móvil, ya que su resolución le permitirá a la aplicación distinguir si el dispositivo está en la sala de estar, en la cocina o en el dormitorio, etc.

Este sistema es de código abierto y totalmente configurable. Es adecuado para aplicaciones domésticas a pequeña escala, así como para aplicaciones empresariales y académicas a gran escala. En el caso pertinente a este proyecto, la maleabilidad del framework probó ser un factor determinante para la elección del mismo como base de trabajo.

A partir de la especificación de este framework, lo que se realizó fue una adaptación del mismo para la aplicación de posicionamiento indoor con la cual se trabajó. Se modificaron así los componentes con los cuales contaba el framework para ajustarlos al problema que se busca resolver.

Estructura del framework

El framework modula el trabajo del sistema en distintos servidores que realizan el trabajo necesario para la localización y generación del mapa de *RSS*. Se divide entonces su funcionamiento en tres partes principales. Durante el proyecto estos

Capítulo 2. Estado del arte

componentes se verían modificados para ajustarse al problema y a mejorar aspectos relativos a la implementación.

Servidor principal Este es el servidor central del sistema, encargado del manejo de flujo de datos y de proveer acceso.

Componentes principales:

- Comunicación entre subsistemas

Controla la comunicación entre la base de datos y el servidor de aprendizaje automático.

- Base de datos

Guarda datos relacionales tales como:

- Historial de ubicaciones de cierto dispositivo.
- Relaciones entre huellas y dispositivos.
- Información sobre los sensores
- Información sobre las distintas ubicaciones.

- Exponer la API del sistema

Esta API es luego utilizada por el Front End para el acceso.

- Exponer el Dashboard del sistema.

Servidor de aprendizaje automático Este servidor se encarga de todo lo relacionado con Machine Learning (ML). El servidor principal envía las huellas subidas por los usuarios al servidor de AA (Aprendizaje Automático) y el mismo se encarga de entrenar el modelo, utilizando varios de los algoritmos discutidos en la sección 2.4.1. Se guarda también en este servidor el estado de los modelos entrenados. La clasificación se lleva a cabo primero separando las huellas totales en un 70 % para entrenamiento de algoritmos y un 30 % para evaluación de los mismos con la métrica de Youden, de forma de asignarles pesos a los algoritmos como se vio en la sección 2.5. Los algoritmos incluidos en la implementación del framework son aquellos listados en la sección de algoritmos determinísticos 2.4.1.

Aplicación de Android para recolección de huellas FIND3 provee una aplicación de Android para el relevamiento de huellas y generación del mapa de RSS. En esta aplicación se etiquetan los datos obtenidos con una ubicación y el dispositivo se encarga de relevar los *RSS* y, mediante la API del sistema, subir la información al servidor principal. La misma aplicación se puede utilizar para llevar a cabo la localización del móvil en una forma de uso básica. La interfaz de la aplicación se puede ver en la figura 2.11. En el campo family se ingresa un nombre que refiere al ambiente general dónde se realiza la experiencia (para poder distinguir, por ejemplo, entre el museo y otro edificio). En el campo device se ingresa el nombre del dispositivo que obtiene las huellas a modo de control.

2.8. Comentarios sobre el estudio del problema

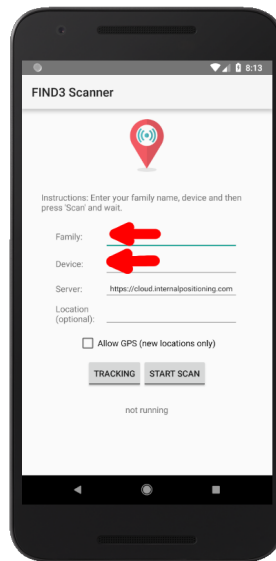


Figura 2.11: Interfaz de aplicación de recolección de huellas de FIND3. Imagen de [25].

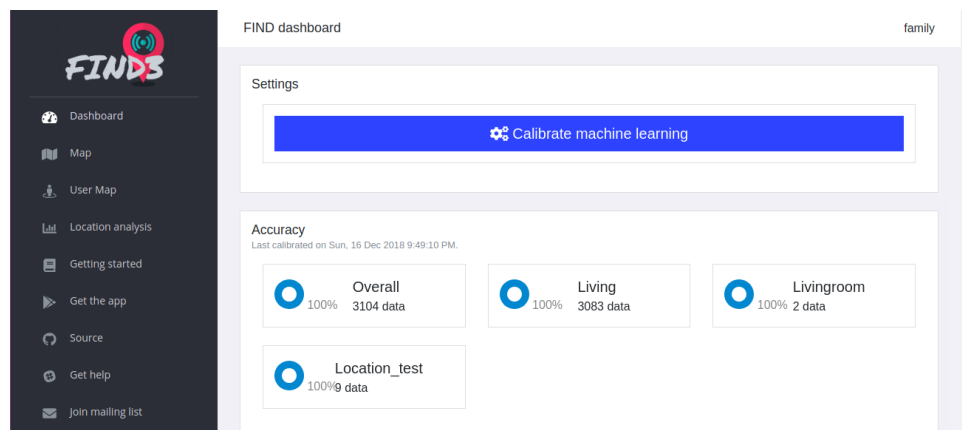


Figura 2.12: Tablero web de FIND3 [25].

Tablero

El sistema de FIND3 cuenta además con un tablero accesible vía web, en el cual se pueden observar fácilmente las huellas subidas al momento, así como la exactitud obtenida sobre el set de huellas de test en el cual se obtuvo la métrica de Youden. Como se ve en la figura 2.12, se puede observar la exactitud del sistema obtenida en el conjunto de prueba para el clasificador.

2.8. Comentarios sobre el estudio del problema

Por último, se puede decir que a partir de este estudio se encontraron los componentes más adecuados para poder tener una solución rápidamente implementa-

Capítulo 2. Estado del arte

ble y que también fuera adecuada para resolver el problema de posicionamiento interior. Contar con un framework que puede ser fácilmente modificado en código, probó ser de gran utilidad, al permitir adaptar el sistema a las necesidades que surgieron en el proyecto de forma sencilla.

Capítulo 3

Implementación de la solución

3.1. Introducción

Luego de la primera etapa donde se estudiaron las distintas tecnologías y posibilidades existentes con respecto al tema, se comenzó la etapa de diseño e implementación del sistema. Las decisiones en cuanto a la tecnología a utilizar y el framework fueron realizadas en una etapa previa, decidiéndose así por el framework FIND3 y uso de señales WLAN con una futura posibilidad de agregar el uso de BLE.

A partir de esto, quedan sentadas las bases sobre las cuales realizar el trabajo de implementación, ajustando lo estudiado al problema en particular en el Museo Nacional de Artes Visuales (MNAV) con los recursos disponibles.

3.2. Pruebas preliminares

En una etapa preliminar al diseño e implementación del sistema, se probó el marco de FIND3 en las oficinas de Tryolabs, empresa en la cual trabaja uno de los integrantes del proyecto. El sistema se levantó utilizando AWS (Amazon Web Services). En esta instancia no se profundizó sobre cambios en la infraestructura para optimizar su rendimiento. El único papel que cumplió este servicio fue de proporcionar un servidor con conexión a internet y una dirección IP pública a donde dirigir la huellas recolectadas por la aplicación de FIND3.

Las oficinas de Tryolabs cuentan con una alta densidad de APs, por lo cual probaron ser un escenario ideal para dar los primeros pasos en la utilización del framework. En particular, se estudió cualitativamente y como experiencia básica dividir el espacio común (la habitación más grande, de aproximadamente 70 m^2) en 4 zonas de tamaño similar. Una vez recolectadas las huellas y entrenado el sistema de clasificación, se logró observar que efectivamente la aplicación realizaba predicciones con una precisión bastante alta (alrededor del 90%). En cuanto a densidad APs, se pudo observar utilizando la aplicación “Wi-Fi Analyzer” [17] que al menos 12 APs eran visibles.

Capítulo 3. Implementación de la solución

A pesar de las diferencias entre el futuro despliegue en el museo que radica en los tamaños de las zonas, la cantidad de huellas y la arquitectura del sistema, se pudo obtener una noción del funcionamiento del sistema y el entrenamiento del mismo previo a realizar la instalación de los APs en el museo.

3.3. Distribución de APs

3.3.1. Diseño inicial

Una vez adquirida la experiencia preliminar con el uso del framework, se comenzó a diseñar la solución específica para el MNAV. Como primer aspecto del diseño inicial, se descartó la posibilidad del uso de infraestructura existente en el MNAV. La misma consistía de 3 APs que no daban cobertura a las zonas de interés. Estos no son suficientes para proporcionar acceso a la red a los visitantes y mucho menos para obtener la densidad necesaria de APs para la localización.

Una de las primeras tareas a realizar fue obtener un estimado de cuántos APs serían necesarios para un sistema de buen rendimiento así como la disposición de los mismos en el museo. Según [25], se pide como mínimo requerimiento para una exactitud aceptable una cantidad mayor a 3 APs cada 50 metros. Sin embargo, al no especificar cuantitativamente esta exactitud se decidió por sobredimensionar la cantidad de APs, teniendo en cuenta que el presente proyecto de investigación prevé analizar cuantos APs son realmente necesarios. Se estableció entonces como requerimiento del proyecto que al menos 5 APs se vieran en todas las zonas del museo, esto independientemente de la conexión que se tuviera. Adicionalmente fue de interés tener una señal entre -65dBm y -70dBm en todas las zonas del museo, para proporcionar conexión básica como un servicio adicional al MNAV.

Para hallar la distribución de APs óptima que cumpliera con los requerimientos se utilizó el software “Ekahau site survey” [12] para planificación de despliegue de redes de Wi-Fi. “Ekahau site survey” es un software que permite diseñar la infraestructura de Wi-Fi de un ambiente a partir del plano, los materiales de la estructura física, el modelo de los APs utilizados y los requerimientos de cobertura que se le exijan. Como forma de sobredimensionar, se consideraron todas las paredes y techos del material hormigón. Los requerimientos de cobertura consisten por un lado en la detección de al menos 5 APs en todo punto del museo y por otro en la conexión básica como servicio al MNAV. Este software generó entonces un reporte con la cantidad y disposición óptimas de los APs. En una primera instancia se llegó a que eran necesarios 15 APs para poder cubrir todo el museo con las especificaciones necesarias.

Luego de generado el reporte inicial con la distribución óptima, se agregaron otras condiciones. Debido a la existencia de 3 APs y para aprovechar las conexiones ya existentes, se decidió colocar 3 APs en sus posiciones. En la figura 3.1 se observa el AP 15 en portería, al AP 14 en secretaría y en la figura 3.2 al AP 1 en biblioteca.

Como consideraciones adicionales, el personal del MNAV pidió contar con un AP dentro de la sala de conferencias y un AP en la sala de área técnica. El AP 13, en la figura 3.1 se dispuso en la sala de conferencias y el AP 6 en la figura 3.2

3.3. Distribución de APs



Figura 3.1: Disposición de los APs en la planta baja del museo simulada en Ekahau.

se dispuso en la sala de área técnica.

Una vez reajustadas las posiciones de algunos APs para cumplir con las últimas condiciones, se volvió a generar el reporte para evaluar el cumplimiento de las condiciones de cobertura. En las figuras 3.3 y 3.4 se mapea el cumplimiento de la visibilidad de APs en la banda de 2.4GHz. Como se puede observar, debido a los cambios de posiciones, algunas áreas cuentan con visibilidad de 4 APs en vez de 5. Este incumplimiento de la condición en el software no se profundizó, debido a que se entendió que hubo un sobredimensionamiento en todo el proceso, por lo que sería suficiente para una buena precisión. En efecto, debido a la configuración de los APs y otros factores descriptos en la sección 3.3.4, se logró visibilidad de al menos 7 APs en todo punto del museo.

Se adjunta a este documento el reporte completo de la planificación generado por Ekahau en el anexo A.

3.3.2. Modelo de APs utilizados

Los 15 APs a utilizar fueron brindados por Antel, y son del modelo Cisco AIR-SAP1602I-A-K9 ilustrados en 3.5. En la hoja de datos del producto [23] se pueden consultar en detalle las especificaciones del mismo.

Capítulo 3. Implementación de la solución

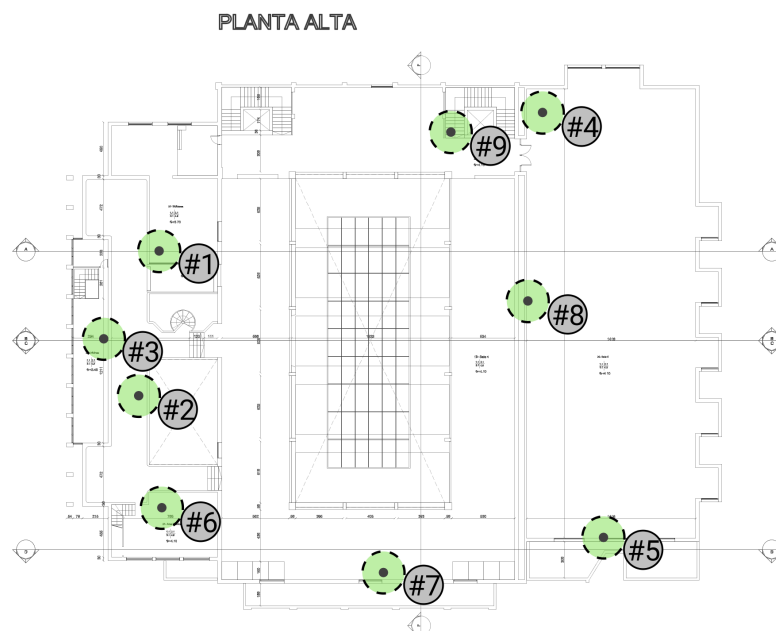


Figura 3.2: Disposición de los APs en la planta alta del museo simulada en Ekahau.

De particular interés es que soporta las bandas de frecuencia en 2,4GHz y en 5GHz (como la mayoría de los APs modernos). Esto significa que cada AP servirá como dos puntos de referencia, debido a la señal asociada a la banda de 2.4GHz y la banda de 5GHz, con una dirección MAC asociada a cada una de estas bandas. El efecto del entrenamiento del sistema con ambas bandas y la localización de dispositivos con el uso de sólo la banda de 2,4GHz se estudia en la sección 4.4.2. Se concluye entre otras cosas que la posibilidad de usar ambas bandas tanto los APs como por los dispositivos permite una mayor precisión.

3.3.3. Configuración de APs

La configuración de los APs fue llevada a cabo con 2 consideraciones básicas. La primera, cumpliendo los requisitos de cobertura y conectividad del MNAV. Esto es, asignar el SSID e IPs fijas dentro de un segmento de red proporcionado. Esta parte, más allá de que se realizó, no es pertinente al proyecto en cuestión y quedó a cargo y responsabilidad del sector de IT del propio museo.

En segundo lugar, se configuraron las potencias y canales en frecuencia a utilizar por los APs. Ambos aspectos están directamente relacionado a las características de la propagación de la señal en el museo. Por lo tanto los canales y las potencias a utilizar por cada AP deben estar fijos para garantizar que el mapa de

3.3. Distribución de APs

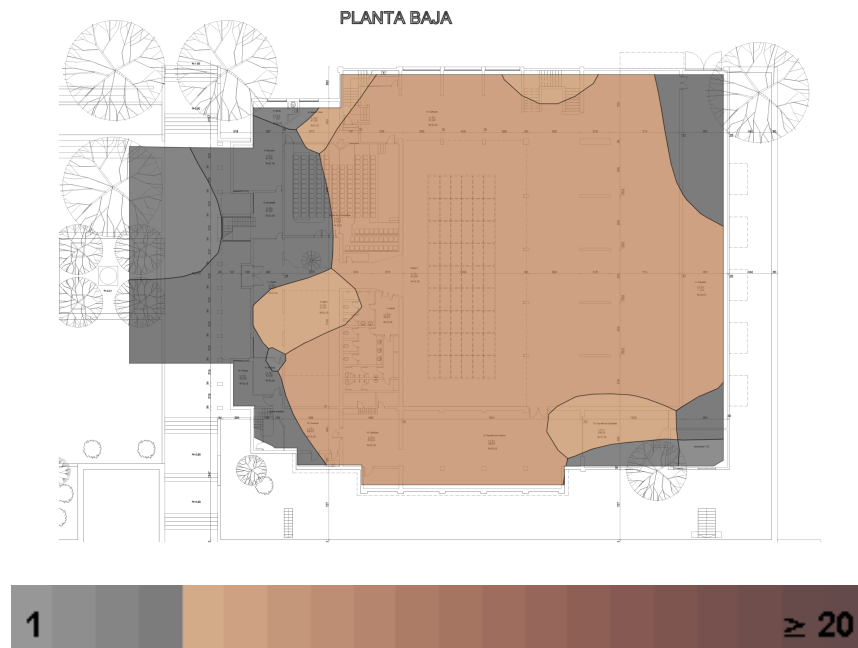


Figura 3.3: Cantidad de APs visibles en cada punto de planta baja en la banda de 2.4GHz. Simulación a través de Ekahau.

RSS no cambie en el tiempo. La planificación de canales se llevó a cabo con la herramienta Ekahau.

3.3.4. Instalación final y resultados

La instalación física de los APs fue realizada por una empresa subcontratada por Antel. Esto consistió en buscar ubicar a los APs a partir de los siguientes factores:

- Ubicaciones cercanas a las planificadas en etapa de diseño con Ekahau
- Considerar la opinión del artista en cuanto a intervenciones que comprometa la estética de la muestra.
- Practicidad en cuanto a mantener el cableado que energiza y da conectividad lo más simple posible.

Se hizo una recorrida con la empresa de instalación y el personal de IT del museo, donde se discutieron estos factores y se fueron definiendo las posiciones finales de los APs.

Capítulo 3. Implementación de la solución

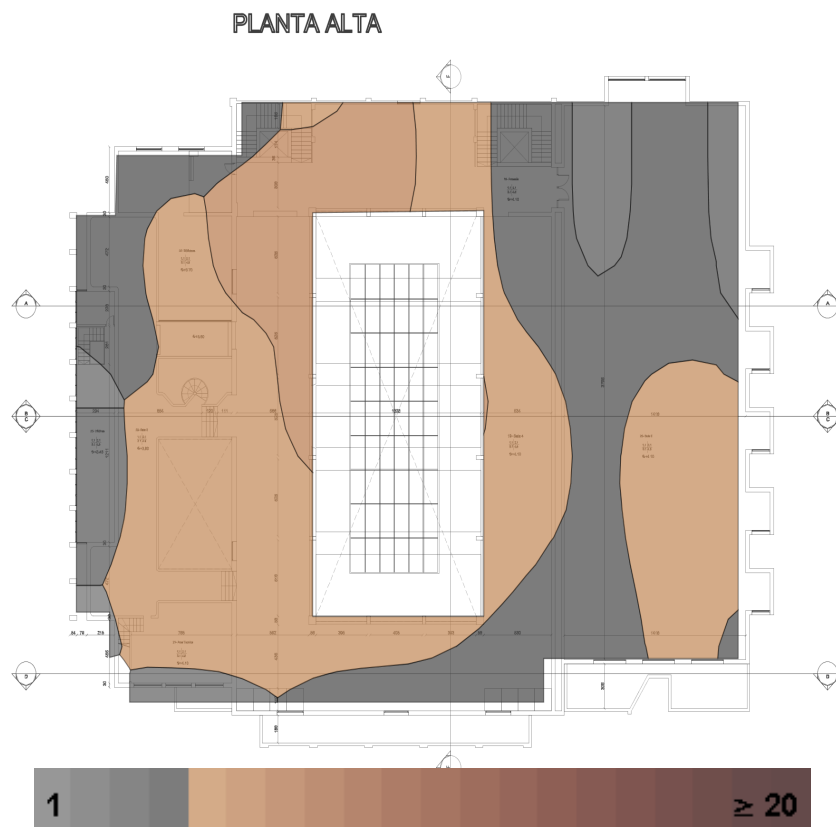


Figura 3.4: Cantidad de APs visibles en cada punto de planta alta en la banda de 2.4GHz. Simulación a través de Ekahau.



Figura 3.5: Modelo de AP utilizado en la instalación del MNAV. Imagen de [23]

3.4. Arquitectura del sistema y modificaciones a FIND3

Por razones prácticas y estéticas de la muestra, la instalación de los APs no se pudo realizar en los lugares específicamente designados en la etapa de diseño. En la figura 3.6 se observan las posiciones finales de los APs en el museo. Se puede observar las diferencias con la planificación realizada con Ekahau en las figuras 3.1 y 3.2.

De todas formas, debido especialmente al sobredimensionamiento en la configuración de la potencia de los APs, se halló que la gran mayoría de los APs son visibles desde cualquier punto del museo. Utilizando la app Wi-Fi Analyzer se fue recorriendo el museo y no se halló ningún lugar donde no se pudieran visualizar al menos 7 APs distintos en la banda de 2.4GHz. Una vez implementado el sistema final, se lograron buenos niveles de precisión. En particular en la sección 4.4.3 se discute el desempeño del sistema si se fueran eliminando APs de la instalación.

De esta forma se concluye que la instalación es satisfactoria en cuanto a garantizar una densidad apropiada de APs.

3.4. Arquitectura del sistema y modificaciones a FIND3

3.4.1. Arquitectura general

Como primera decisión respecto al despliegue del sistema, se decidió por utilizar AWS (Amazon Web Services) [20]. Esta plataforma permite un fácil, ágil y efectivo despliegue, así como distintos servidores que se ajustan a las funciones necesarias para el proyecto. Más allá de eso, el grupo de trabajo cuenta con experiencia en el trabajo con AWS, lo cual probó ser muy valioso para un rápido y correcto despliegue del sistema.

La arquitectura final del sistema se decidió entonces en base a la utilización de AWS y distintas consideraciones para el proyecto y de eficiencia general. En la figura 3.7, se pueden observar los distintos componentes e interacciones. Los componentes de la arquitectura son los siguientes:

- Servidor principal
- Servidor de AA (aprendizaje automático)
- Servidor de información
- Base de datos
- Almacenamiento de AA (aprendizaje automático)
- Almacenamiento de media (material multimedia)

A continuación se describen las distintas adiciones y modificaciones al sistema de FIND3 por cada componente.

Capítulo 3. Implementación de la solución

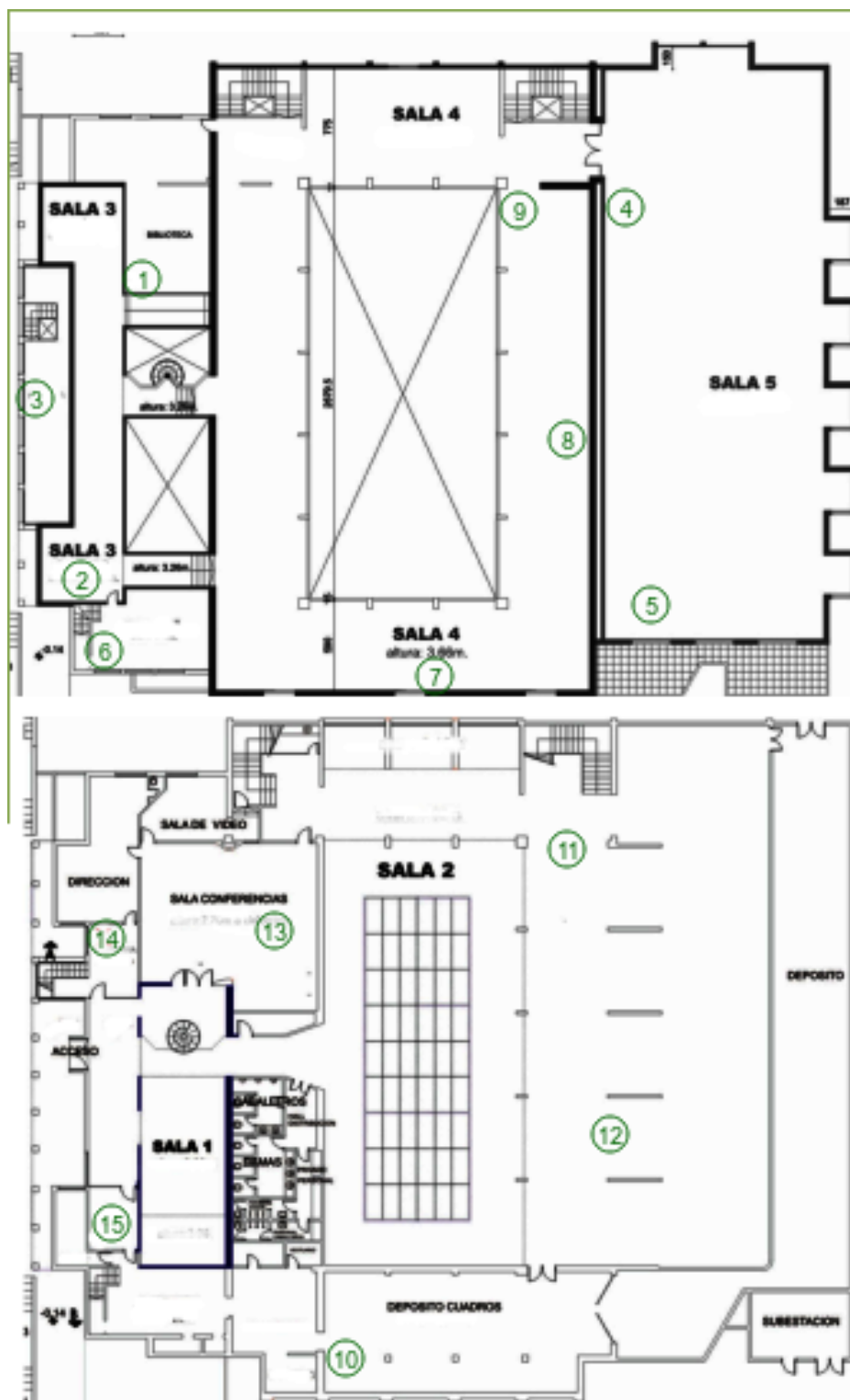


Figura 3.6: Distribución final de APs luego de la instalación.

3.4. Arquitectura del sistema y modificaciones a FIND3

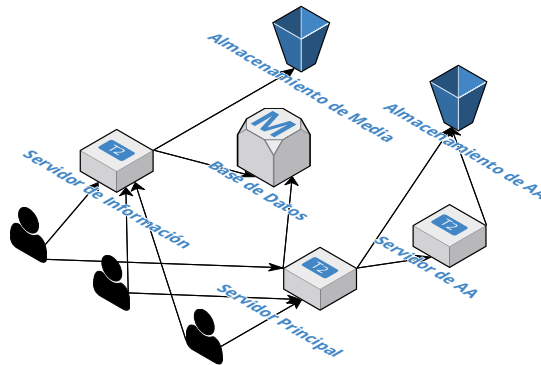


Figura 3.7: Arquitectura del sistema implementado a partir de FIND3

3.4.2. Componentes de la arquitectura

- Servidor principal

El servidor central del sistema se encuentra en un contenedor de Docker, en una instancia EC2 (Amazon Elastic Compute Cloud) t2 (tier 2) de tamaño medio (2 vCPUs, 2.3 GHz, Intel Broadwell E5-2686v4, 4 GiB de RAM).

EC2 es un servicio que permite a sus usuarios comprar servidores virtuales, pagando por hora utilizada. Cuenta con muchos tipos de instancias, dependiendo de las necesidades de cada usuario.

El servidor principal es el encargado de recibir todos los request hacia el backend, procesarlos y derivarlos al Servidor de AA si es necesario. Esto incluye el dashboard de FIND3, donde se ve el estado actual del sistema. El servidor principal se comunica con el Servidor de AA, con el Almacenamiento de AA y con la Base de Datos.

Se cambió toda la capa de comunicación con la base de datos de FIND3, ya que el mismo utiliza SQLite. La razón de este cambio es debido a la escalabilidad, SQLite no escala con varios usuarios accediendo simultáneamente, y se busca usar este sistema en un ambiente con múltiples usuarios. Por esta razón se añadió la base de datos como un ente separado en otro servidor.

- Servidor de aprendizaje automático (AA)

Este servidor está en la misma EC2 que el Servidor Principal, pero en otro contenedor de Docker.

Este servidor se encarga de entrenar y clasificar el modelo cada vez que el

Capítulo 3. Implementación de la solución

Servidor Principal lo solicite. Este servidor atiende comunicaciones exclusivamente del Servidor Principal y se comunica solamente con él mismo y el Almacenamiento de aprendizaje automático.

Como único cambio respecto a FIND3, se guarda la metadata (estado de los algoritmos) así como los CSV (comma separated values) de huellas en el Almacenamiento de aprendizaje automático.

- Servidor de información

Está levantado en una EC2 distinta, de tipo t2 micro (1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB de RAM). Se encarga de proveer información (texto, audio e imágenes) sobre las obras que hay en determinada zona. En la base de datos están registradas las distintas obras, con su respectivo texto y urls a las imágenes y audio.

Este componente es totalmente nuevo y solamente es accedido desde la App de Android. El mismo solamente se comunica con la Base de Datos.

- Base de datos

Es una base de datos almacenada en RDS (Amazon Relational Database Service). RDS es un servicio de Base de Datos Relacionales que ofrece AWS, es muy similar a EC2, pero con Bases de Datos.

Este componente es totalmente nuevo y cumple las funciones de base de datos que solían estar comprendidas por el Servidor Principal.

- Almacenamiento de aprendizaje automático (AA)

Es un bucket de S3 (Simple Storage Service). Esto es una API que provee almacenamiento de manera escalable, fiable y rápida (especialmente se trabaja intra-AWS) donde se guarda información del Servidor de AA.

Aquí se almacena el estado de los algoritmos del modelo de AA, así como un CSV que contiene los datos con los que se entrenó al modelo. Este componente también fue creado en los cambios.

- Almacenamiento de material multimedia (media)

Es también un bucket de S3. Este contiene las imágenes y audios de todas las obras. Este componente también es nuevo.

3.4.3. Interacciones entre los componentes

A continuación se describen las diferentes interacciones entre los componentes del sistema.

Flujo de re-entrenamiento del modelo de AA

El flujo de re-entrenamiento refiere al proceso necesario para entrenar al sistema, en particular al relevarse nuevas huellas.

3.4. Arquitectura del sistema y modificaciones a FIND3

- Se hace un GET Servidor Principal, a la url (*/calibrate*).
- El Servidor Principal se comunica con la Base de Datos para conseguir todos los datos etiquetados, que se usan para entrenar el modelo.
- Separa los datos en distintos sets (70 % de entrenamiento y 30 % de verificación).
- Crea un CSV con los datos de entrenamiento y los sube al Almacenamiento de AA.
- El Servidor Principal llama al Servidor de AA, pasando como parámetro la url al CSV para que re-entrene.
- El Servidor de AA se comunica con el Almacenamiento de AA, levantando el estado de todos los algoritmos utilizados y el CSV para el entrenamiento.
- El Servidor de AA entrena todos los modelos.
- El Servidor de AA le avisa al Servidor Principal que dicha tarea finalizó.
- El Servidor Principal se encarga de evaluar los datos de evaluación, comunicándose con el Servidor de AA.
- El Servidor Principal se encarga de calcular los índices de Youden.

Localización de una nueva huella

El proceso de localización de un dispositivo, requiere en una etapa previa enviar al sistema los datos de RSSI en forma de una huella sin etiquetar. El sistema procede entonces a clasificarla, guardando los resultados para este dispositivo en la base de datos.

- Se hace un POST al Servidor Principal, a la url */data*, pasando la huella sin etiquetar en el body de request.
- El Servidor Principal guarda en la Base de Datos dicho request.
- El Servidor Principal le envía la huella al Servidor de AA.
- El Servidor de AA clasifica la huella con todos los algoritmos del modelo.
- El Servidor de AA envía resultados al Servidor Principal.
- El Servidor General aplica los índices de Youden, obteniendo así el resultado final.
- El Servidor Principal guarda dicho resultado en la Base de Datos.

Capítulo 3. Implementación de la solución

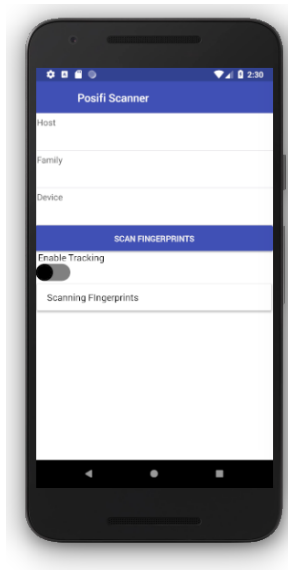


Figura 3.8: Interfaz de aplicación de recolección de huellas nuevo.

Localización de un dispositivo

Una vez clasificada la huella enviada por el dispositivo, el resultado de la clasificación se encuentra en la base de datos del sistema. La localización consiste en obtener esta información.

- Se hace un GET al Servidor Principal, a la url */location/posifi/id_dispositivo* (id de dispositivo es un identificador único de cada dispositivo que se comunica con el sistema).
- El Servidor Principal se comunica con la Base de Datos para conseguir la ultima estimación de la localización del dispositivo.
- El Servidor Principal retorna dicha información.

3.4.4. Aplicación de recolección de huellas

Luego de probar la aplicación de recolección de huellas de FIND3, se pudo observar que en los dispositivos que tienen Android el escaneo Wi-Fi depende esencialmente del fabricante del hardware, junto con algunas configuraciones de usuario y otras aplicaciones que se están ejecutando (otras aplicaciones pueden solicitar el escaneo de redes Wi-Fi). Esto implica que en la práctica no hay un “tiempo de escaneo estándar” para Wi-Fi. Esto dificultaba la posibilidad de tener un tiempo estimado del escaneo de las huellas. Por lo tanto, se implementó una aplicación de recolección de huellas propia, que sustituye a la de FIND3.

La aplicación se realizó utilizando el framework React Native. React Native [13] permite crear aplicaciones móviles usando solo JavaScript. Se construye una aplicación móvil real que es indistinguible de una aplicación construida usando

3.5. Distribución de zonas de localización

Java. React Native utiliza los mismos componentes básicos de la interfaz de usuario que las aplicaciones normales de Android. Sólo se deben colocar esos bloques de construcción juntos usando JavaScript y React. Se decidió utilizar este framework debido principalmente a la familiaridad del grupo de proyecto con JavaScript.

Básicamente, se reescribió la aplicación usando React Native, teniendo en mente que se deben poder modificar los tiempos de consulta sobre las redes que el dispositivo detecta en el momento de la consulta. De esta forma, se busca poder tener una experiencia de recolección de huellas mas rápida y confiable.

3.4.5. Corrección en utilización de huellas

En una etapa de análisis de datos posterior a la implementación, se pudo observar que no se estaban utilizando todos los datos guardados en la base de datos para entrenar los algoritmos. Este error en el sistema fue detectado posteriormente a la instalación del sistema en el MNAV, ya que estaba implementado dentro del framework en el módulo encargado de cargar los datos a un csv. En particular, el sistema contemplaba un máximo de 1000 huellas, mientras que en realidad se habían recolectado para entrenar 10469. Esto es de particular gravedad no sólo por estar usando tan sólo un 10 % de los datos, sino que la distribución de como elegir este 10 % de datos era realizada al azar. Apenas detectado este problema, se lo corrigió tanto en un ambiente de pruebas como luego en producción. Luego de hacer esta primera optimización, se pudo ver una mejora de casi un 5 % en el sistema, se pasó de un promedio de 90 % de exactitud a un 95 %. Esto significa en la práctica una reducción de la cantidad de errores en un 50 %.

Luego de haber agregado todos los datos, el tiempo de entrenamiento aumentó de forma no sustentable. A pesar de que el tiempo de entrenamiento no es tan sensible como el tiempo de clasificación, los tiempos excesivos de entrenamiento dificultaban los cambios o re-entrenamientos sobre el sistema en despliegue. Por eso fue que se decidieron eliminar aquellos algoritmos cuyo tiempo de entrenamiento aumentó demasiado a causa de la mayor cantidad de datos. Como el orden de SVM no lineal es $O(N^3)$ para algunos kernel que se estaban utilizando, con un aumento considerable en los datos este algoritmo generaba complicaciones para recolectar nuevas huellas. Una vez eliminado SVM no lineal, el tiempo de entrenamiento bajó a niveles aceptables, por lo cual se decidió no eliminar otro algoritmo bajo este criterio.

3.5. Distribución de zonas de localización

Una vez definida la arquitectura del sistema y la instalación de los APs, se abordó la implementación de la localización específicamente para el museo. Un aspecto central al diseño de la localización es establecer las zonas que el sistema podrá discernir. Esto es, en una localización discreta hay una cantidad finita de zonas que pueden resultar de la localización del usuario. Para esta tarea se tomó en cuenta como factor principal la experiencia final del usuario, discutiendo en

Capítulo 3. Implementación de la solución

términos subjetivos y probando con la app final la mejor distribución de zonas para ello.

3.5.1. Consideraciones realizadas

El primer factor a decidir concierne a la resolución espacial del sistema. A partir de lo estudiado de sistemas de localización utilizando Wi-Fi, las experiencias de pruebas preliminares y la densidad de obras distintas por área del museo, se determinó que haya una relación de más de una obra por zona. Esto condujo también a la conclusión que buscar discernir entre muchas zonas de pequeño tamaño traería consigo mayor especificidad pero menor exactitud, lo cual se traduciría en una peor experiencia de usuario dada que la especificidad pocas veces sería de mucho valor.

Como segundo factor principal, se tomaron de base los diseños de los modos de funcionalidad descritas en las secciones 3.6.3 y 3.6.4. Estos diseños, descritos en la sección 3.6, fueron realizados en conjunto con el área educacional del museo, teniendo en cuenta la experiencia del usuario final. En términos de impacto sobre las decisiones de zonas, se resalta que la modalidad descrita en la sección 3.6.4, requiere de la capacidad de diferenciar entre zonas cercanas a 8 obras específicas en la planta baja. Estas obras por lo tanto deberán pertenecer necesariamente a zonas diferentes. Por ejemplo en la figura 3.9, se ilustra la distribución de zonas finales en la planta baja. Cada una de estas zonas cuenta con una de las obras centrales a la modalidad “Recorrido a ciegas”.

Por último, se priorizó una simplicidad que permitiera no solo el correcto despliegue a tiempo del sistema sino una mejor capacidad de probar la experiencia y detectar falencias en la localización.

3.5.2. Relevamiento de primeras huellas

Una vez tomadas estas consideraciones, se fue iterando sobre distintos diseños de distribución de zonas. De esta forma se fueron recolectando huellas de RSSI y etiquetándolas con las zonas marcadas.

El relevamiento de huellas se realizó utilizando siempre el mismo dispositivo para mantener uniformidad. El dispositivo utilizado fue el celular de uno de los integrantes del grupo, un Motorola modelo moto g6. Este cuenta con soporte para las dos bandas de frecuencia (2.4GHz y 5GHz). El relevamiento se realizó recorriendo cada zona con la aplicación de recolección de huellas. La disposición del dispositivo con respecto al usuario fue de forma de emular lo más posible a un futuro usuario de la aplicación recorriendo el museo mientras mira la pantalla del celular. Esto es, sujetado en una mano cerca del cuerpo y dispuesto con la pantalla apuntando hacia arriba a nivel del pecho. Se buscó realizar estas medidas con distintas orientaciones del usuario, para mitigar el efecto de la orientación de la caracterización de las zonas.

Durante este proceso de relevamiento de huellas y re-entrenamientos, se procedía a probar exhaustivamente la experiencia del usuario con la app desarrollada

3.6. Aplicación de Android: MNAVegante

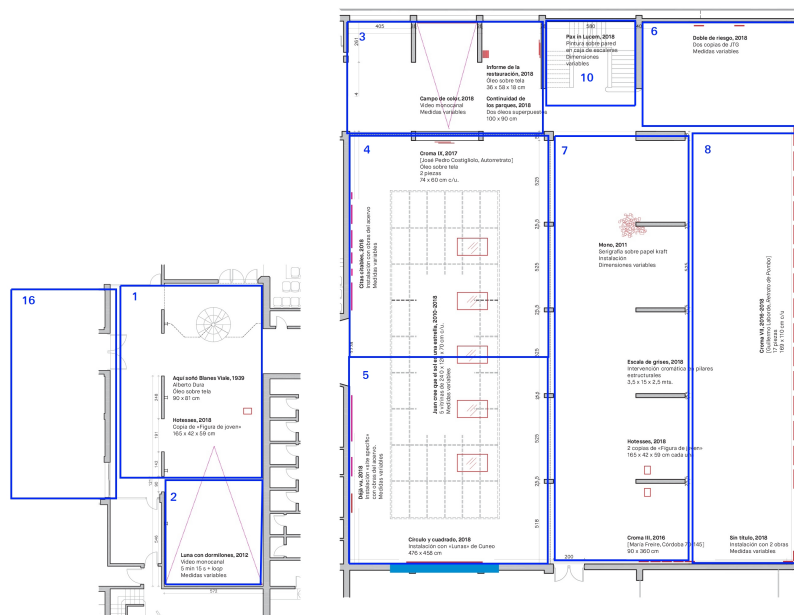


Figura 3.9: Distribución de zonas en planta baja del museo. La zona 16 incluye el exterior del mismo. En rojo se muestra la distribución de las obras.

en paralelo. Esto permitió ir detectando las zonas más conflictivas a medida que se realizaban distintos diseños. El resultado final en términos de distribución de zonas se puede observar en los planos de las figuras 3.9, 3.10 donde se demarcan las 16 zonas utilizadas.

El conjunto de huellas final consistió en un total de 10469 huellas, distribuidas entre las 16 zonas acorde al tamaño de las mismas. De esta forma, la zona 7 tuvo la mayor cantidad con 1096 huellas mientras que la zona 2 fue la de menor cantidad con tan solo 357 huellas.

3.6. Aplicación de Android: MNAVegante

Se realizó una aplicación en Android para la navegación de los visitantes al MNAV. La aplicación para la navegación dentro del museo se basó en la aplicación open source del museo de Craft and Folk [2] discutida en la sección 2.7.

La aplicación busca hacer uso de la localización indoor a través del sistema implementado. A través de ella, se busca enriquecer la experiencia del recorrido del museo en dos modalidades, proporcionando acceso a contenido sobre la muestra

Capítulo 3. Implementación de la solución

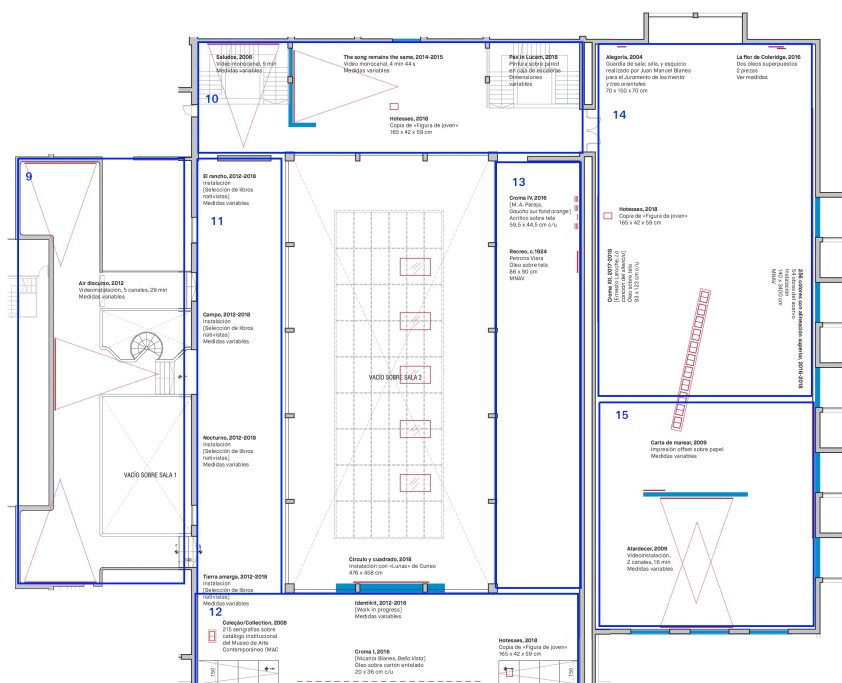


Figura 3.10: Distribución de zonas en planta alta del museo. En rojo se muestra la distribución de las obras.

de una forma interactiva y en función de la posición del usuario. En un primer acceso a la aplicación se despliega la pantalla inicial mostrada en 3.11a, donde se presenta la muestra y sus partícipes. En la captura de pantalla del menú principal de la figura 3.11b se despliegan los tres modos de uso.

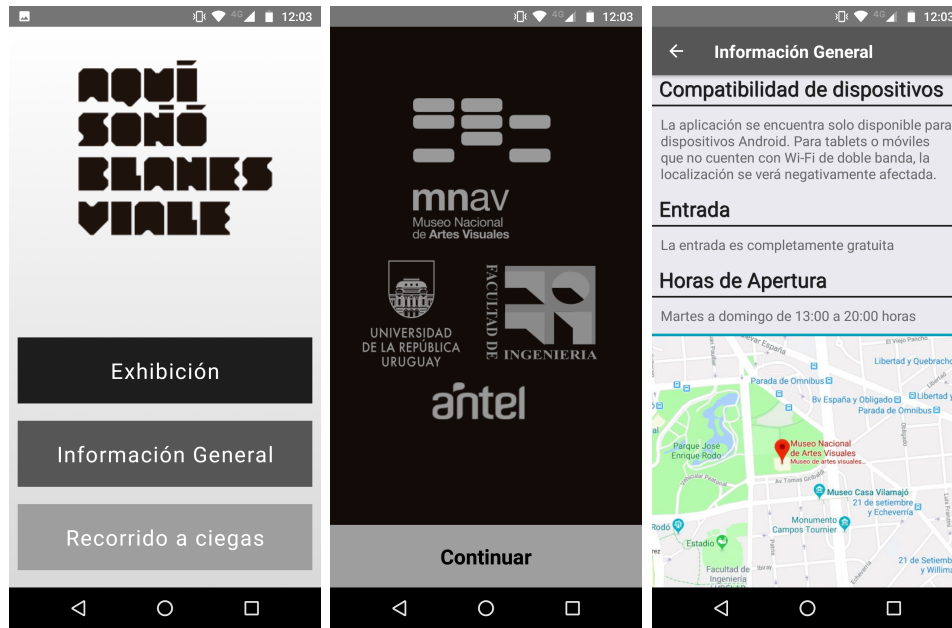
MNAVegante se encuentra disponible en la Play Store, donde puede ser descargada de forma gratuita por cualquier dispositivo con versión de sistema operativo Android 4.1 o posterior.

3.6.1. Interacción entre aplicación y sistema de localización

El primer aspecto que se trabajó sobre la aplicación es su forma de acceder al servicio de localización y al contenido de la muestra. A continuación se describe el flujo general de esta interacción entre aplicación y sistema, que luego es usada por las funciones de Exhibición y Recorrido a ciegas. Los pasos realizados por el sistema entre estas interacciones se pueden hallar en la sección 3.4.3.

- El dispositivo mide los valores de RSSI de los APs circundantes como una huella sin etiqueta.

3.6. Aplicación de Android: MNAVegante



(a) Pantalla inicial. (b) Menú principal. (c) Información general.

Figura 3.11: Pantallas iniciales y menús principales de la aplicación MNAVegante.

- Hace un POST al Servidor Principal a la url `/data`.
- Hace un GET al Servidor Principal a la url `/location/posifi/id_dispositivo`
- Con el resultado obtenido calcula si debe cambiar la zona actual. Este calculo depende de si se encuentra en Exhibición o en Recorrido a Ciegas y se describe en más detalle en las secciones 3.6.3 y 3.6.4.
- En caso de que no sea necesario el cambio de zona, termina el bucle.
- En caso de que sea necesario, la app se comunica con el Servidor de Información para obtener información de la nueva zona.
- El Servidor de Información busca dicha información en la Base de Datos y la devuelve a la app.
- La app busca las imágenes y audios necesarios del Almacenamiento de Media.
- La app utiliza esta información de acuerdo a la interacción del usuario como se detalla en las secciones 3.6.3 y 3.6.4.

Las modalidades de Exhibición y Recorrido a ciegas hacen uso de la localización indoor para estimar constantemente donde se encuentra el usuario durante su recorrido. En la práctica, esto se traduce a un par de POST y GET de http, en los cuales se obtiene la localización actual. Este proceso se lleva a cabo una vez por segundo.

Capítulo 3. Implementación de la solución

Para un efectivo uso de esta localización de forma que se traduzca a una buena experiencia de usuario, se debe diseñar una forma de obtener robustez suficiente en cuanto a exactitud complementada con una fluidez propia de una aplicación dinámica.

El uso de la aplicación en una etapa de prueba probó ser muy útil en cuanto a identificar los parámetros más importantes en cuanto a experiencia de usuario, ya que es difícil cuantificar en una primera instancia el impacto de los mismos sin visualizar el resultado final. Dadas las distintas características entre Exhibición y Recorrido a ciegas la utilización de la localización se hizo de distintas maneras.

3.6.2. Información general

En el menú de información general, visto en la figura 3.11c, se busca proporcionar la información básica de la app, la muestra y el museo independientemente de la conectividad del usuario. Actualmente es la única sección funcional fuera del museo, ya que no hace uso del posicionamiento logrado a través de los APs ubicados dentro del museo. En ella se incluyen los siguientes ítems:

- Aquí soñó Blanes Viale: Descripción breve de la muestra
- Funcionamiento del recorrido de la exhibición: Explicación de modalidad “Exhibición”
- Recorrido a ciegas: Explicación de modalidad “Recorrido a ciegas”
- Compatibilidad de dispositivos: Descripción de dispositivos para los cuales la app está disponible (Android) y hacen uso de doble banda de Wi-Fi (aquellos solo con 2.4GHz se verán negativamente afectados en cuanto a localización como se discute más adelante en la sección 4.4.2).
- Entrada: Expresar que la entrada es gratuita
- Horas de Apertura
- Mapa con ubicación del MNAV

3.6.3. Exhibición

La modalidad Exhibición busca proporcionar al usuario acceso al contenido referente a las obras cercanas a su ubicación. De esta forma se da un servicio adicional al usuario.

Desde el punto de vista del usuario, se le irán desplegando las obras cercanas a medida que recorra el museo como se observa en la figura 3.12. Estas se irán actualizando automáticamente, cambiando dependiendo de la zona detectada por el móvil a través de la localización.

En caso de que el usuario quiera obtener información sobre la obra, puede acceder a la misma utilizando el botón en la parte inferior de la imagen. De esta forma, accede a un menú fijo (no cambia hasta que el usuario salga hacia atrás) donde se despliegan un texto y un audio descriptivo de la obra.

3.6. Aplicación de Android: MNAVegante

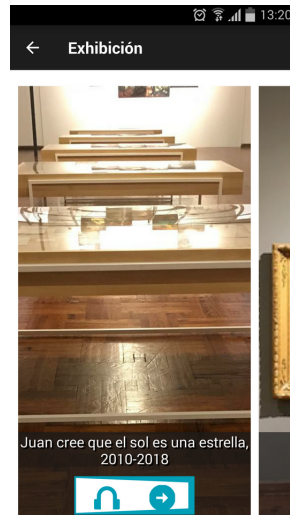


Figura 3.12: Modo exhibición. La captura de pantalla despliega distintas imágenes con información correspondiente a las obras cercanas. Entrando mediante el botón inferior, se puede obtener información específica de la obra.



Figura 3.13: Modo exhibición. Información desplegada de la obra. Incluye texto descriptivo, la imagen y un audio reproducible donde se describe la obra.

Capítulo 3. Implementación de la solución

Tiempo de actualización de pantalla

Más allá de que el sistema se relocaliza una vez por segundo como fue discutido en la sección 3.6.1, el saltar de grupo de obras en grupo de obras constantemente es molesto para el usuario. Por otro lado, el aumentar el tiempo entre actualizar las obras desplegadas permite luego utilizar métodos de acumulación de intentos de localización, lo cual permite tener una mayor seguridad al momento de desplegar las obras. Esto tiene un límite principalmente en cuanto a que si se aumenta demasiado, la pantalla tarda demasiado en cambiar entre zonas. Adicionalmente, más allá de que se busca evitar un error entre zonas, en los pocos casos donde esto se da (ya sea por estar en zonas borde o por contar con una ráfaga de errores), la pantalla se quedaría trancada mucho tiempo en este error antes de poder ser solucionado.

Criterio de selección

A pesar de que el sistema proporciona la localización, dado que existe un tiempo entre actualización de pantallas, es importante aprovecharlo para ganar seguridad sobre la posición del usuario. En una primera aproximación, solo se cambiaba la zona cuando se recibía 3 veces seguidas una zona con una probabilidad mayor al 30 %.

A medida que se fueron probando distintas ideas simples en el front-end se terminó diseñando un buffer circular que almacenara las probabilidades de todas las zonas. Este buffer consiste en las probabilidades de que el usuario se halle en cada zona para cada uno de los últimos 5 intentos de localización (por lo tanto, 5 segundos). Esto permite hacer un mayor uso de la información que el de solo la localización adivinada y además genera una inercia para evitar cambios abruptos que solían darse por errores en ciertas posiciones. Luego de sumar las probabilidades almacenadas en todo el buffer, la aplicación cargará el contenido de la zona con mayor suma de probabilidades. En la práctica, el efecto de usar el buffer fue obtener una mayor robustez en las medidas así como cierta inercia para agregar estabilidad a las predicciones obtenidas. Esto fue estudiado desde la experiencia del usuario, ya que el análisis de precisión del sistema del capítulo 4 no toma en cuenta las herramientas de la aplicación.

3.6.4. Recorrido a ciegas

Motivación

El recorrido a ciegas parte de una propuesta concreta, donde se busca acercar la muestra y el museo en general a un público con capacidad visual reducida. La intención de esta modalidad es proporcionar una experiencia interactiva que dependa lo menos posible de imágenes o acciones con la pantalla del dispositivo. Esto permitiría no sólo el acceso a audio descriptivo de las obras alrededor sino una independencia mayor al momento de usarla.

El acceso a contenido de la muestra en esta modalidad corresponde a 8 obras ubicadas en planta baja del MNAV. El objetivo concreto sería entonces diseñar una

3.6. Aplicación de Android: MNAVegante

forma por la cual a medida que se recorriera las instalaciones del MNAV, el sistema identificara la proximidad del usuario a alguna de estas obras y se comenzara a reproducir el audio respectivo a la misma.

Requerimientos de solución

Para empezar se plantean las principales características necesarias del sistema.

- Robustez de exactitud:

En comparación a la modalidad exhibición, la robustez del sistema cobra una mayor relevancia. Es preferible que el sistema tarde en decidirse cuando está en duda a que se comience a reproducir un audio equivocado, ya que estos tienen una longitud importante y un usuario podría quedar atascado con el audio incorrecto y sin la obra descrita en su cercanía.

- Recorrido flexible:

El usuario debe poder recorrer las instalaciones en cualquier orden, por lo que no hay un orden preestablecido de obras a reproducirse.

- Evitar repetición:

Debido a que un usuario puede pasar cierto tiempo alrededor de una obra o puede volver a pasar por allí sin intención de volver a escuchar el audio de la misma, es necesario que los audios no sean repetibles.

- Manejo simple y sin necesidad de botones:

El usuario debe poder utilizar la aplicación sin mayores dificultades de complejidad y no debe precisar hacer uso de botones o equivalentes.

Funcionamiento final

La solución propuesta parte de que cada obra se encuentra en una zona de localización distinta. De esta forma, la proximidad a una obra se tomará como equivalente a encontrarse dentro de la misma zona que esta obra, por lo que no se hace uso de un recorrido fijo. Tal como en la modalidad Exhibición, el dispositivo se encuentra realizando consulta al sistema una vez por segundo, obteniendo la predicción de localización en el proceso.

Una vez que se ingresa a la modalidad Recorrido a ciegas, se despliega una imagen genérica del museo y se reproduce un primer audio instructivo. Luego de finalizado el audio comienza a ser utilizada la localización, sin ser necesaria ninguna futura interacción con botones o similar.

Para asegurar robustez de exactitud, se consideró necesario que el resultado de la localización resulte en la misma zona tres veces consecutivas para considerar al usuario en dicha zona. Una vez establecido que el usuario se encuentra en proximidad de una obra, se comienza a reproducir un audio correspondiente a la obra y se despliega una imagen de la obra con el título de la misma. Más allá de no

Capítulo 3. Implementación de la solución

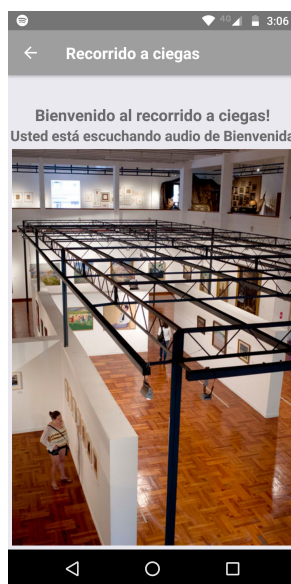


Figura 3.14: Modo Recorrido a ciegas. Se despliega una imagen de la obra detectada, en este caso es una imagen genérica para marcar el inicio del recorrido. Se reproduce un audio inicial instructivo de cómo utilizar la aplicación.

depender de la capacidad visual del usuario, tener una referencia básica de como se ve la obra puede ser útil para usuarios que cuentan con algo de capacidad visual.

Otro factor que fue de utilidad fue la falta de repetición. El sistema no permite que una vez en el recorrido a ciegas se detecte la misma zona dos veces hasta que se salga de la modalidad. Esto es de utilidad no solo para evitar repetición, sino que el ir descartando zonas a medida que el usuario las recorre agrega una mayor robustez a la exactitud, evitando posibles confusiones con zonas ya recorridas. La introducción de la falta de repetición se dio en las primeras pruebas y resultó en uno de los mayores saltos de desempeño, debido a que eliminó la mayor cantidad de errores de Recorrido a ciegas.

3.6.5. Aplicación web para información de museo

Para poder facilitar el trabajo de tener la información del museo necesaria para desplegar en MNAVegante, se realizó una aplicación web. Se creó en React, una biblioteca JavaScript para crear interfaces de usuario y de esta forma recolectar la información de las obras. La aplicación web se despliega como en las figuras 3.15 y 3.16.

Esta aplicación consiste simplemente de un dashboard para obtener las imágenes, descripciones y demás información sobre las obras del museo. Los campos en la figura 3.16 reflejan la información necesaria, siendo Posifi ID el identificador de en que zona se encuentra la obra y “Es del recorrido para ciegos?” determina si se usará para Exhibición o para Recorrido a ciegas. En el caso de que una obra se encuentre en más de una zona, se deberán ingresar dos entradas idénticas con distintos Posifi ID.

3.7. Inauguración de la muestra

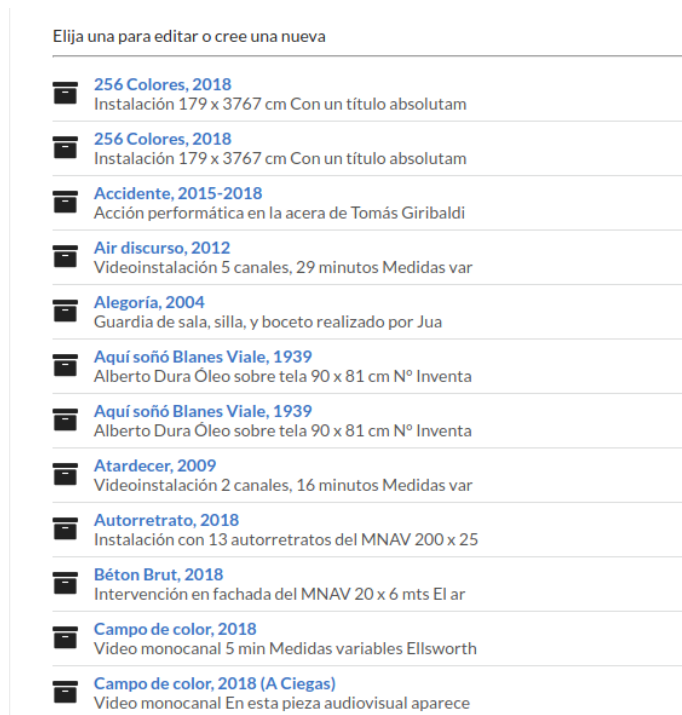


Figura 3.15: Menú principal de aplicación web para información de la muestra. Aquí se carga la información de todas las obras a ser cargadas luego por la aplicación MNAVegante.

3.7. Inauguración de la muestra

La etapa principal de implementación concluyó formalmente en la inauguración de la muestra Aquí soñó Blanes Viale el 8 de noviembre de 2018. En la inauguración el sistema se probó en condiciones anómalas debido a que el museo se encontró lleno de visitantes. A pesar de que estas condiciones pudieran tener un impacto en la conectividad y en la propagación de las señales de Wi-Fi, la aplicación tuvo un buen funcionamiento. Las críticas de la aplicación en su gran mayoría fueron positivas, con la excepción de observaciones respecto al pobre rendimiento para dispositivos sin 5GHz o respecto a un bug donde la localización falla y la pantalla queda en blanco. El primer fenómeno se estudia en detalle en la sección 4.4.2, mientras que el bug de la aplicación presentó dificultades para ser replicado. Lo que sí se logró acotar con respecto al bug es que no depende del dispositivo o su sistema operativo, pudiendo un dispositivo eliminar el bug al desinstalar y reinstalar la aplicación. Con respecto a bloqueos (salidas abruptas de la aplicación), MNAVegante presentó algunos problemas, especialmente con aquellas tablets que no cuentan con acceso a internet más allá del Wi-Fi del museo. Estos bloqueos se pudieron corroborar que eran debidos a pérdidas de conectividad a internet y se suponen directamente relacionados a alguna de las bibliotecas nativas de Android utilizadas para acceder al contenido de audios o imágenes del servidor.

Los únicos cambios realizados sobre el sistema posteriores a la inauguración de la muestra fueron realizados en consecuencia a un error encontrado en el framework

Capítulo 3. Implementación de la solución

Pagina para subir informacion de Obras
Complete los campos que quiera editar y haga click en editar informacion

Usted esta editando la obra : 256 Colores, 2018

Obra	256 Colores, 2018
------	-------------------

Posifi ID	location_14
-----------	-------------

Instalación
179 x 3767 cm

Con un título absolutamente descriptivo se vuelve sobre el tema de la unicidad

☐ Es del recorrido para ciegos?

Imagenes...

 Elegir Imagenes

Audios...

 Elegir Audio

Editar Informacion

Eliminar

Atras

Figura 3.16: Campos para edición de información sobre cada obra en la aplicación web.

por el cual se utilizaba solo un 10 % de las huellas, descrito en la sección 3.4.5.

Capítulo 4

Análisis de desempeño

4.1. Introducción

Una vez finalizada la implementación inicial del sistema y de lograr asegurar una buena experiencia de usuario, se buscó realizar un estudio amplio con respecto a las posibilidades actuales y futuras del sistema.

Para medir la precisión del sistema se utilizó el mismo código que se encuentra funcionando en la nube, replicado en un ambiente de pruebas Jupyter Notebooks. Los Jupyter Notebooks [14] son documentos producidos por la aplicación Jupyter Notebook App, que contienen tanto código informático (por ejemplo, python) como elementos de texto enriquecido (párrafo, ecuaciones, figuras, enlaces, etc...). Los Notebooks son tanto documentos legibles por la persona, conteniendo la descripción del análisis y los resultados (figuras, tablas, etc.), como documentos ejecutables para realizar análisis de datos.

En primer lugar se trasladaron todos los datos que estaban presentes en la base de datos a un archivo con formato csv, formato de datos separados por coma, para así poder cargarlos utilizando la biblioteca csv de Python. Estos datos se recolectaron con las medidas de RSS del Wi-Fi realizadas en la etapa de relevamiento de huellas. Cada huella en el archivo csv cuenta con una columna para la etiqueta (la zona real en la cuál fue tomada) y luego columnas que corresponden a la potencia recibida para cada AP. Estos datos fueron anonimizados y están disponibles para el uso público de los mismos en la url: https://github.com/ffedee7/posifi_mnav/tree/master/data_analysis.

Luego de tener los datos disponibles, se procedió a definir una metodología de forma tal que se pudiera evaluar el rendimiento del sistema bajo los mismos criterios. De esta forma también se distinguieron los conjuntos de entrenamiento, prueba y validación.

La primera prueba realizada consistió en probar el sistema actual en su totalidad en cuanto a su rendimiento. De esta forma se pudo tener un mayor entendimiento del sistema y ver cuáles son las limitantes, ventajas y desventajas del mismo.

Luego se observó el comportamiento del sistema al no detectar a un AP en una

Capítulo 4. Análisis de desempeño

medida de RSS. El framework rellena estos valores como 0dBm. Se analizó el impacto de reemplazarlo por el valor de -90dBm como medida más fiel a estar alejado de un AP. Como siguiente parte del análisis total, se tomaron decisiones en cuanto a omitir algoritmos que resultaron redundantes o directamente contraproducentes a la localización.

Como primer análisis de posibles cambios, se estudiaran los datos teniendo en cuenta el efecto de los APs en el rendimiento. Para esto se alteraron los datos recolectados para emular los distintos escenarios posibles. Primeramente se observó el efecto de filtrar las huellas acorde a las MACs visibles. Esto es, filtrando los datos utilizando solo aquellas medidas de potencia provenientes de APs específicos o incluso en una banda de frecuencia específica. En una primera instancia, se observó el efecto de utilizar solamente los APs instalados y configurados comparándolo con el sistema actual que utiliza todos los APs observados en la recolección de datos. Luego, se estudió el rendimiento del sistema en un dispositivo que no soporta doble banda de Wi-Fi, evaluando a su vez posibles causas de las diferencias y proponiendo una alternativa para mitigar problemas. Por último, esta herramienta permitió observar la precisión del sistema acorde a que APs no estén presentes en el entrenamiento del mismo o en la clasificación de un dispositivo. La falta de APs en el entrenamiento permite inferir cuántos APs son necesarios para generar un sistema de cierta precisión, mientras que la falta en clasificación permite observar la robustez del sistema a partir de posibles fallas en APs que queden fuera de servicio.

Siguiendo en la línea de cantidad de APs que son necesarios instalar para alcanzar distintas precisiones objetivo, se hizo un análisis de la cantidad de huellas necesarias para obtener cierta precisión. Esta prueba junto a la de cantidad de APs de entrenamiento sirven de forma de analizar los costos incurridos a la hora de instalar APs y horas de trabajo recolectando huellas en función de la precisión que garantizan.

Otro aspecto analizado en cuanto a lo realizado en la etapa de diseño inicial consistió en evaluar la distribución de zonas de localización. En esta prueba se levantó un sistema paralelo dónde se obtuvieron datos con medidas de RSS para zonas de menor tamaño y luego se analizó la evolución de la precisión a medida que se juntan para obtener zonas más grandes. Esto da una idea de la resolución espacial que se puede lograr con el sistema actual así como las posibles metodologías para distribuir las zonas.

Por último, se realizó un estudio básico a partir de los datos recabados de los recorridos realizados con la aplicación. Estos datos relevados durante la extensión de la muestra incluyen las zonas más visitadas así como la hora a la cuál fueron visitadas.

4.2. Metodología y nomenclatura utilizada

Se eligieron tres estadísticas principales para estudiar el rendimiento del sistema, que tienen sentido para la aplicación.

La primera medida que se eligió utilizar para poder medir el rendimiento del

4.2. Metodología y nomenclatura utilizada

sistema es la precisión definida como en la sección 2.5.2. Esta métrica es importante para el análisis porque representa en porcentaje cuántas de las predicciones realizadas como zona x provenían realmente de una huella de la zona x . Esto se traduce en cuanto a la experiencia de usuario en que si el usuario obtiene una predicción, en la se encuentra en una zona de precisión alta, es altamente probable de que efectivamente se encuentre en ella. La precisión se puede entender entonces como la utilidad de la predicción una vez realizada. Por lo cual, esta métrica promediada entre todas las zonas pareció ser la más adecuada para estimar el rendimiento del sistema.

Como segunda métrica se decidió utilizar la exhaustividad, que se define de la siguiente forma:

$$Exhaustividad = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Negativos}.$$

La exhaustividad permite representar cuántas huellas que son efectivamente de la zona x son finalmente clasificadas en esa zona. Es decir, si un usuario se encuentra en una zona en concreto, cuál es la probabilidad de que el sistema lo clasifique dentro de esa zona. Esta métrica en cierta medida permite complementar la precisión y da una buena idea de en que zonas los usuarios pueden tener mayores problemas para ser localizados.

La tercera métrica utilizada fue la matriz de confusión, que permite reconocer la confusión entre las distintas zonas. Para la posición de la fila i y columna j en la matriz, el valor representa cuántas huellas correspondientes a la zona i fueron clasificadas en la zona j . De esta forma, en un buen clasificador la diagonal de la matriz debe tener los números más elevados, dado que serían los casos donde la clasificación fue correcta. Cuando existen huellas clasificadas en una zona errónea, se puede estudiar si las huellas de una zona suelen ser clasificadas en alguna zona adyacente en común. Este conocimiento permite un mejor accionar para en principio evitar o al menos mitigar los posibles errores a lo que esto conduce. Para tener una noción de lo que significa en la distribución espacial de las zonas referirse a las imágenes 3.9 y 3.10 donde se presenta la numeración de las zonas del museo.

Por otro lado, para tener una mejor estimación de los resultados se hizo un promedio de haber ejecutado las predicciones del sistema 10 veces, debido a que en cada ejecución la distribución de los datos en los distintos conjuntos (entrenamiento, prueba y validación) se realizaba en forma aleatoria.

4.2.1. Criterio de aceptación del sistema

Con el fin de tener un criterio para poder definir cuantitativamente un funcionamiento deseable del sistema se definieron ciertos valores de referencia para la precisión. La bibliografía utilizada sobre el tema no desarrolla sobre aplicaciones concretas y la experiencia cualitativa del usuario. Por lo tanto, no se pudo encontrar una definición acertada de cuál es el valor óptimo para poder evaluar el buen desempeño del sistema. Es por esto que se decidió definir este valor de forma subjetiva acorde a lo que se pudo juzgar en experiencia de usuario con diversas

Capítulo 4. Análisis de desempeño

precisiones promedio. Luego de haber transitado varios ajustes de los parámetros y pruebas del sistema se decidió que el punto de inflexión donde el sistema puede decirse que ya no tiene un rendimiento aceptable es en 90 % de precisión promediando todas las zonas. Para la precisión por zonas individuales se fue un poco mas laxos, obteniendo de la experiencia un valor de 80 % como valor aceptable. Estas precisiones de referencia coinciden con las obtenida utilizando tan solo el 10 % de las huellas, dónde el sistema funcionaba aunque con una baja cantidad de errores en las predicciones. Como punto de referencia para un buen rendimiento, se considera el 95 % obtenido luego de incluir todas las huellas. En este escenario, los errores de clasificación fueron muy escasos, permitiendo una experiencia más fluida.

Es a partir de estos valores que de aquí en adelante se evaluará el rendimiento del sistema frente a todas las variables posibles.

4.2.2. Conjuntos de entrenamiento, prueba y validación

Para poder obtener las medidas de desempeño del sistema se dividieron los datos en los siguientes tres conjuntos definidos de la siguiente manera:

- Conjunto de entrenamiento: Subconjunto de los datos utilizados para entrenar a cada modelo.
- Conjunto de prueba: Subconjunto de datos utilizado para evaluar a los modelos entrenados, también utilizado para calcular la estadística de Youden la que ayuda a elegir los pesos de los distintos algoritmos.
- Conjunto de validación: Subconjunto de datos utilizado para validar el sistema luego de haber entrenado y poder asegurar que se tiene un modelo que representa bien la realidad. Este es el conjunto a partir del cuál se infiere el rendimiento real del sistema.

Recordando que el framework FIND3 realiza una división de los datos de 70 % para entrenamiento y 30 % para la estadística de Youden. El mismo también utiliza estos para la validación de los algoritmos. Esto en la practica no es una buena medida dado que se esta utilizando los mismos datos con los que se entrenó. Es por esto que se decidió modificar esta división para hacer lugar al conjunto de datos de validación con el cual se validaron las distintas estadísticas para el desempeño del sistema. Para esto, en lugar de separar los datos en un 70 % de entrenamiento y 30 % de prueba, se separaron en un 70 % de entrenamiento, 20 % de prueba y 10 % de validación. La separación de estos conjuntos fue realizada al azar dentro de cada etiqueta pero manteniendo la proporción de huellas con respecto al total.

4.2.3. Omisión de algoritmos

Como primer acercamiento al estudio de desempeño se experimentó con distintas combinaciones de los algoritmos. Rápidamente se detectó que tres de los nueve algoritmos no aportaban valor a la hora de la clasificación. Esto es, a partir de

4.3. Primer acercamiento al problema

realizar el entrenamiento con y sin esos algoritmos, en todos los intentos se obtuvo un rendimiento mejor o igual dejando afuera estos.

De esta forma se decidió eliminarlos a modo de simplificar el sistema. Los algoritmos que se eliminaron fueron QDA y los dos algoritmos de Naive Bayes. En el caso de QDA se debió a su pobre rendimiento en todas las pruebas realizadas, siendo que quitar el algoritmo nunca bajó la precisión. En el caso de los algoritmos de Naive Bayes, la decisión se tomó dado que las implementaciones del mismo contenían algunos errores de programación. Estos algoritmos fueron implementados por los creadores del framework, a diferencia de los anteriores algoritmos implementados desde la biblioteca de machine learning scikit-learn [21], por lo cual los resultados obtenidos con los mismos no eran los esperados.

Los algoritmos finalmente utilizados para el resto de las pruebas fueron los siguientes:

- SVM lineal
- Decision trees
- Random forests
- MLP
- AdaBoost
- KNN

Luego de haber visto estos resultados, los análisis posteriores fueron hechos utilizando únicamente estos algoritmos.

4.3. Primer acercamiento al problema

Como primera prueba, se utilizaron todas las huellas relevadas y se realizaron las pruebas sin modificaciones significativas al sistema. La precisión obtenida en promedio fue de 95,7%.

Cabe destacar que estamos hablando de un resultado en promedio ya que para las distintas pruebas se realizaron varias tiradas con las que luego se realizó un promedio. Esto es válido desde aquí en adelante cuando se hable de un resultado en promedio.

En la figura 4.1 se observa la matriz de confusión obtenida para el total de los datos. Se observa que la diagonal cuenta con la gran mayoría de las huellas de prueba lo que representa que estamos en presencia de un buen clasificador como lo detalla la figura 4.2.

En cuanto a las confusiones que se pueden observar los mayores problemas se presentan entre las zonas adyacentes 6, 7 y 8. Por ejemplo, 10 huellas de la zona 8 fueron clasificadas como 6 o 7 y 16 huellas de las zonas 6 y 7 fueron clasificadas como zona 8. También se observan confusiones más aisladas pero también entre zonas adyacentes, como la 1 y la 2. Esto se condice con la práctica, donde se

Capítulo 4. Análisis de desempeño

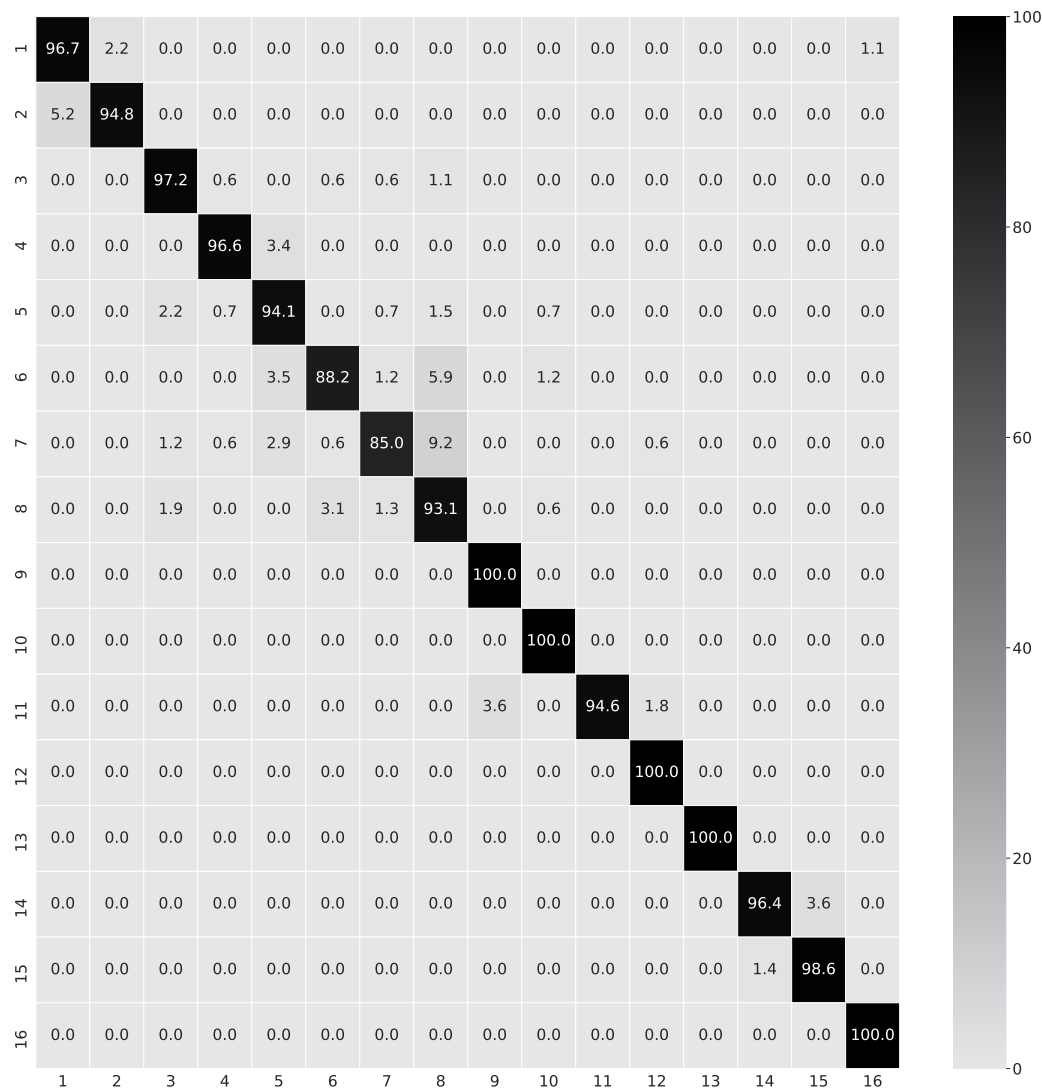


Figura 4.1: Matriz de confusión para el total de los datos

observaron las mayores problemáticas de confusión en las zonas que poseen mayor cantidad de zonas adyacentes. Otra de las hipótesis de la confusión de estas zonas fue que las mismas estaban en una zona con un ascensor. Es posible que el ascensor genere interferencias en la señal, lo que se traduce luego en tener huellas disimiles debido a este fenómeno afectando entonces la clasificación de las mismas.

En la figura 4.2 se puede observar la precisión y exhaustividad de forma granular por zona. En ella podemos ver que tenemos en la mayoría de las zonas una medida aceptable de precisión. Sin embargo, las zonas en las que matriz de confusión aparecen con mayor índice de confusión presentan menor valor de exhaustividad como es de esperarse.

4.3. Primer acercamiento al problema

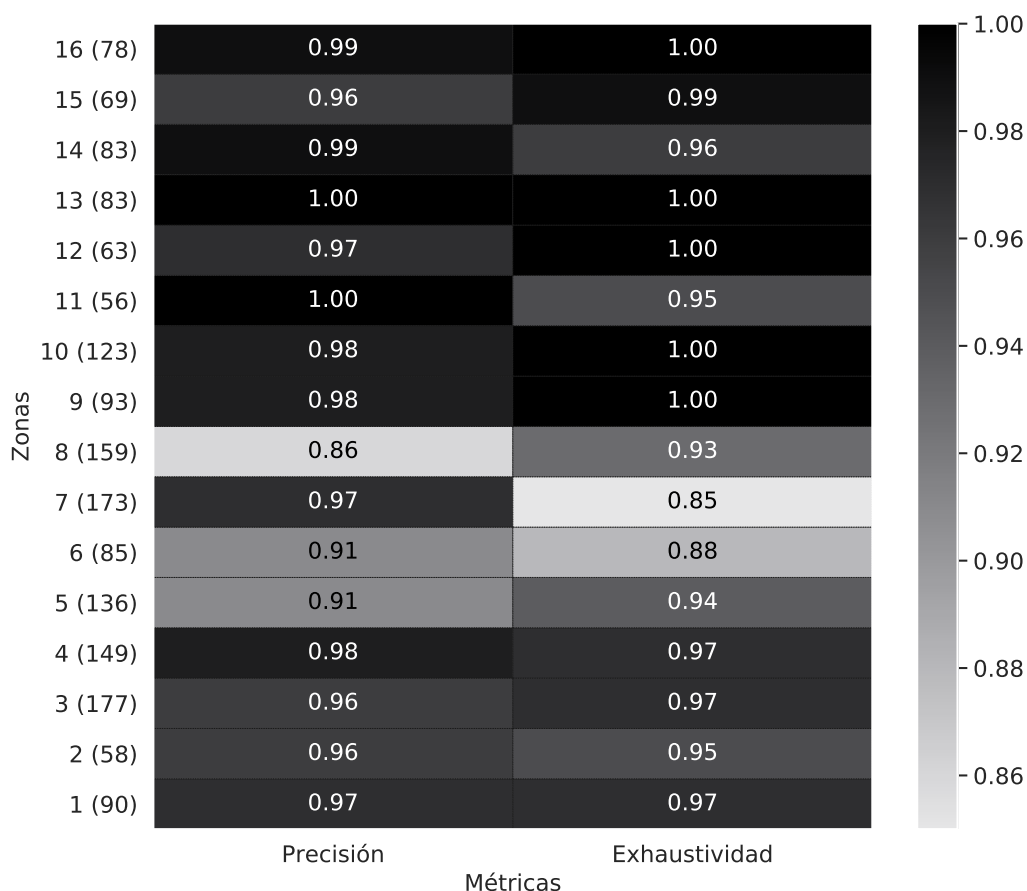


Figura 4.2: Tabla de exhaustividad y precisión para todos los datos. Entre paréntesis se anota la cantidad de muestras para cada zona.

4.3.1. Representación en potencia de APs no detectados

Se pudo observar que dentro de los datos que se levantaron de la base de datos habían datos faltantes (es decir, no todas las columnas del csv poseen un valor definido). Esto es de esperarse, ya que en un tiempo dado un dispositivo no capta la señal de RSS de todos los APs que el sistema tiene registrado. Luego de hacer esta observación, se investigó el impacto de esto en la representación de los datos hacia los algoritmos de machine learning, ya que las librerías que los implementan no trabajan con datos faltantes. Se pudo observar que el sistema implementado en FIND3 reemplaza los datos faltantes con el valor 0dBm. En primera instancia se ejecutaron las distintas pruebas para las diferentes estadísticas con estos valores. Luego se buscó un valor que tuviera más sentido para la aplicación. Como se sabe que un valor indefinido representa que para esa muestra el AP no está siendo visto por el móvil, entonces fue intuitivo buscar un valor que en vez de ser cero se asemeje al valor de un AP que se encuentra muy lejano. Por ello se probó el valor de -90dBm como una mejor aproximación de tener un AP a una distancia lejana.

Resultados obtenidos:

Capítulo 4. Análisis de desempeño

- 95,7 % para 0dBm
- 96,8 % para -90dBm

Se observa que se aumentó en un 1 % la precisión con la presentación de los APs no detectados con -90dBm. Este 1 % es relevante debido a la alta precisión, representando alrededor de 25 % de los errores del sistema anterior. Sin embargo, en la práctica esta diferencia no se hizo notar, ya que la variabilidad en la aleatoriedad de la elección de los conjuntos presentó un mayor impacto. Otro factor a tener en cuenta es el margen de error de la estimación que es aproximadamente un 3 % con lo cual estos valores caen dentro del mismo intervalo.

Otro factor a tener en cuenta es que al estar frente a un problema de clasificación el número adjudicado a los valores faltantes es indiferente dado que el algoritmo aprende independientemente del significado de la aplicación. Es decir, siempre que el valor esté fuera del rango de trabajo de RSS, los algoritmos tenderán a aprender rápidamente que esas entradas deben ser ignoradas.

4.4. Entrenamiento con selección de MACs

En esta sección se probaron dos limitantes que pueden surgir a la hora de la puesta en producción del sistema. Una de ellas es el impacto de variar la cantidad de APs con el cuál el sistema es entrenado. La otra es el análisis del impacto de uso de diferentes bandas de frecuencia. Es aquí donde el filtrado de datos según las direcciones MAC interviene.

Se pudieron filtrar los datos según las direcciones MAC de los APs para que no pertenecieran al conjunto de muestras en etapa de entrenamiento o en etapa de clasificación. Para realizar este filtrado se eliminaron aquellas columnas del conjunto de datos que representaran la MAC asociada. Dado que cada dirección MAC corresponde a la banda de 2,4GHz o de 5GHz de un único AP, se pudo estudiar el impacto de cada banda de frecuencia. Eliminando las MACs pertenecientes a la banda de 5GHz se pudieron probar los efectos de la falta de una banda. Luego en otra prueba aparte se eliminaron una cantidad arbitraria de APs filtrando las MACs de los mismos en las dos bandas.

4.4.1. Filtrado según APs instalados

En primer lugar, se observó que los datos del sistema son tales que las huellas no solo contienen los APs instalados y configurados en el museo sino que también poseen información de otros APs circundantes. Para realizar un estudio más acotado de la situación se decidió restringir los datos a los APs conocidos. Esto es útil para ver si se pierde precisión utilizando esta menor cantidad de APs con tal de mantener el control sobre las señales utilizadas por el dispositivo para localizarse.

La figura 4.3 muestra los resultados en la matriz de confusión. Se observa de esta figura que los resultados en cuanto a confusiones de las zonas son muy similares, siguiendo el mismo patrón que siguen los datos con todos los APs presentes.

4.4. Entrenamiento con selección de MACs

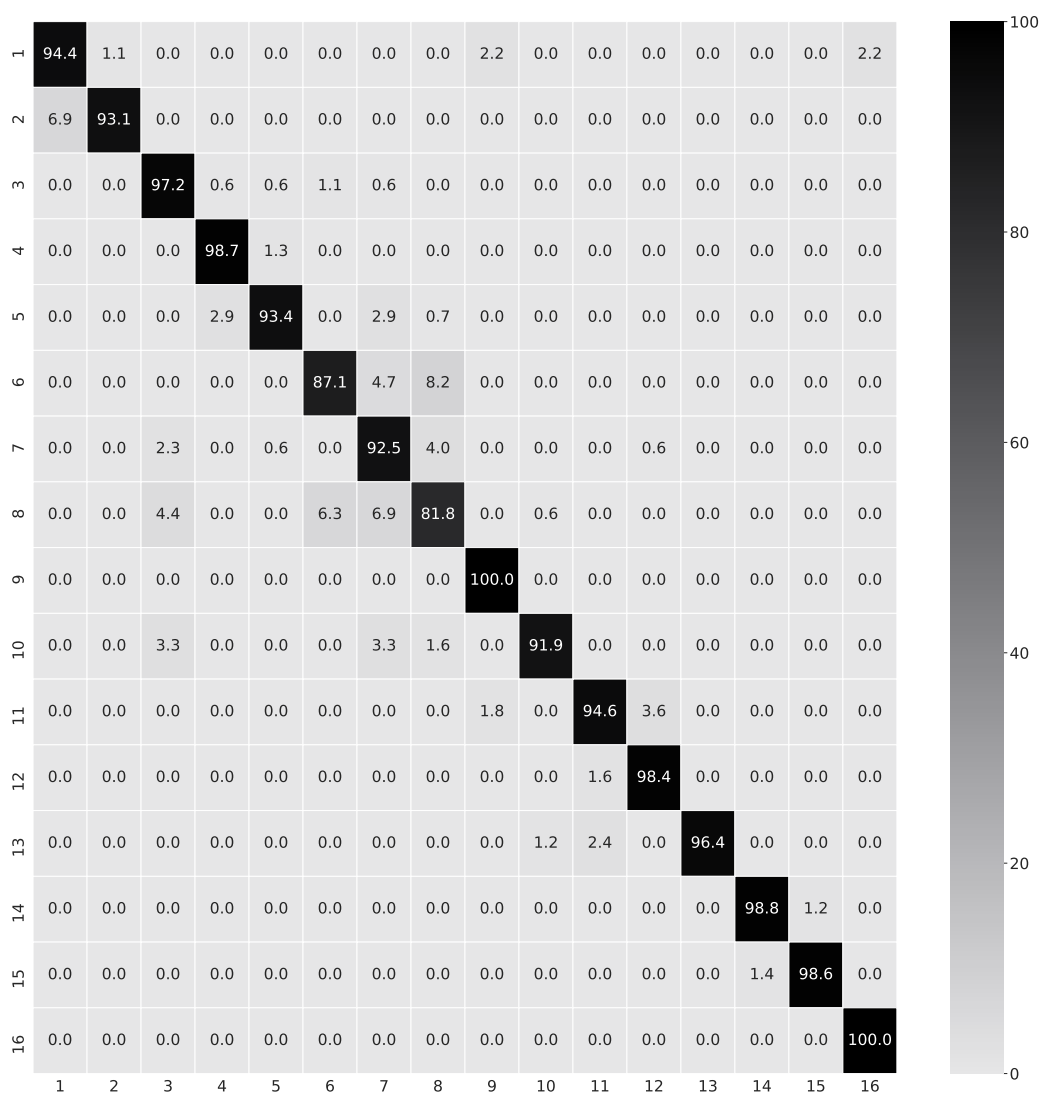


Figura 4.3: Matriz de confusión para las MACs conocidas

Esto entonces refleja un cambio mínimo en cuanto a usar la suma de aquellos APs instalados más los APs circundantes frente a usar solamente los APs del diseño del sistema. Una de las conclusiones a las que se puede llegar es que se hizo una buena estimación de la cantidad de APs que se necesitaban para poder tener una buena precisión dentro del museo, esto junto a que se hizo una buena distribución de las zonas contribuyó a que se pueda obtener una excelente experiencia dentro del museo. Por lo tanto, no es necesario tener la información de los APs que no son instalados o configurados específicamente para la localización.

En cuanto a precisión se obtuvo un 95,6 %, que no difiere mucho del resultado que se había obtenido con todos los APs del sistema. Esto confirma el análisis anterior que pone en perspectiva la baja importancia de los APs no instalados en comparación a la incertidumbre que conllevan. Además, se debe remarcar que el

Capítulo 4. Análisis de desempeño

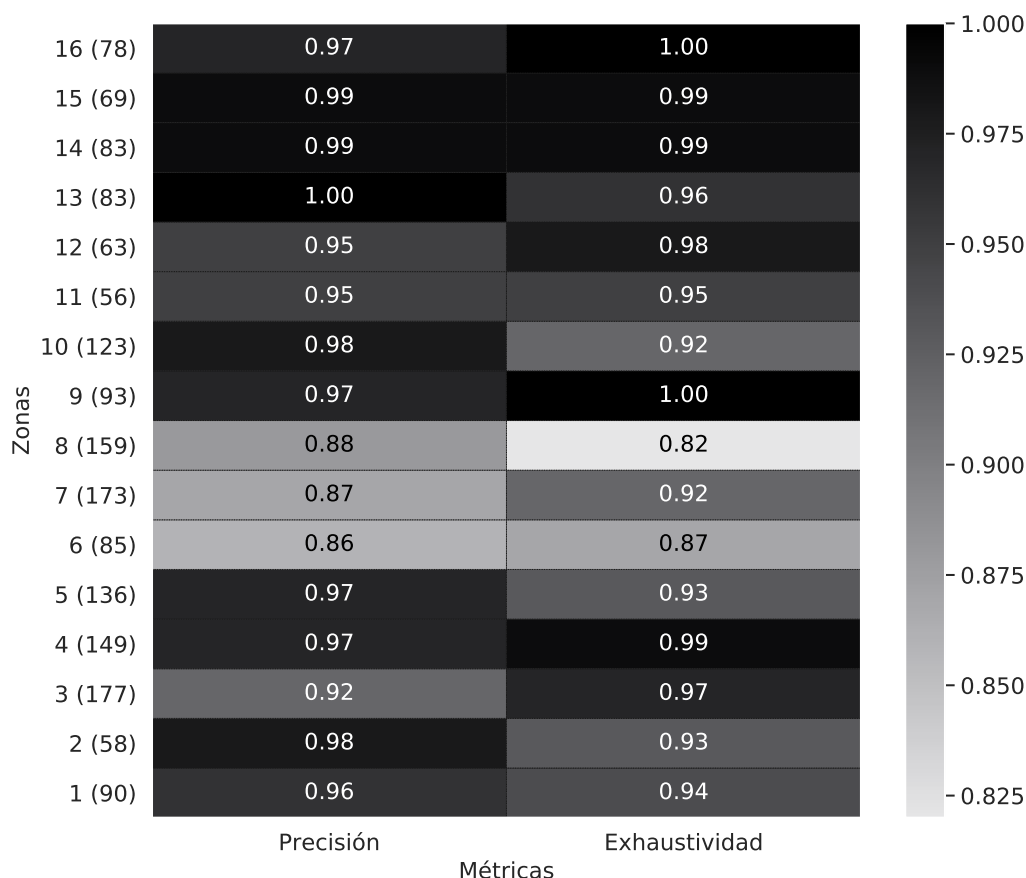


Figura 4.4: Tabla de cobertura y precisión de las zonas para MACs conocidas

análisis de precisión se lleva a cabo con huellas tomadas en un mismo intervalo de tiempo corto, a diferencia de las clasificaciones realizadas por usuarios dos meses luego de empezada la muestra. Por lo tanto, se concluye que trabajar con los APs seleccionados es una buena forma de añadir robustez al sistema.

Se puede decir entonces que es preferible trabajar con los APs instalados, filtrando por ellos y sin aceptar la posible detección de APs adicionales en el entrenamiento o localización. Posiblemente si el sistema permanece mucho tiempo, los APs que no se pueden controlar cambien y afecten el sistema. Teniendo control sobre cada AP utilizado se puede evitar un deterioro futuro de la precisión.

Por ultimo vemos en la figura 4.4 que la precisión por zona esta dentro de los valores aceptados y que la exhaustividad que se ven más afectadas son aquellas que se confunden más.

4.4.2. Entrenamiento para dispositivos de una banda

En la actualidad existen dispositivos que cuentan con tarjeta de red que soporta doble banda (5GHz y 2.4GHz) pero también existen aquellos que no cuentan con esta característica. Es importante entonces tener un análisis de las diferen-

4.4. Entrenamiento con selección de MACs

cias entre ambas configuraciones. A partir de esto es que se decidió analizar el rendimiento para aquellos dispositivos con una sola banda. La hipótesis a probar es que al realizar el entrenamiento con un dispositivo que trabaje con las bandas de 2.4GHz y 5GHz, una medida de potencia que no incluya señales de 5GHz será clasificada erróneamente debido a que la clasificación no tiene en cuenta la posible diferencia entre móviles. Para probar esto primero se entrenaron con todos los MACs conocidos, los de 5GHz y los de 2.4GHz, para luego clasificar con huellas que tuvieran solamente 2.4GHz.

Como se ve en la imagen 4.5 se puede observar que la clasificación de estas huellas es realmente mala. Hay muchas zonas confundidas y por esto muchas veces no se pueden distinguir entre las distintas zonas. Con esto podemos comprobar la hipótesis de que los celulares de una banda tienen un mal funcionamiento cuando se clasifican con el sistema que entrenó con todas las huellas. El comportamiento reflejado en la matriz de confusión 4.5 coincide con la experiencia de utilizar la aplicación con un dispositivo sin doble banda. En las experiencias observadas, el dispositivo permanentemente se ubica en las zonas 6, 10 y 16 independientemente de la posición. Esto evidencia que ciertas características de estas zonas particulares se asemejan a un escenario dónde no se reciban señales de 5GHz, por ejemplo siendo zonas con menor densidad de APs cercanos.

El parámetro de la precisión dio solamente un máximo de 40 % que esto otra vez nos indica el mal funcionamiento dado que está muy por debajo del criterio de aceptación.

Se observó entonces que una de las mayores carencias de la aplicación final fue el rendimiento de la misma para dispositivos que solo cuentan con la banda de 2.4GHz.

Para poder mitigar este problema se estudió la posibilidad de tener un sistema en paralelo entrenado solamente con huellas que tengan en cuenta los APs de 2.4GHz. Esto permitiría que el dispositivo que recolectó huellas y el clasificado se encuentren en igualdad de condiciones, evitando el problema antes mencionado.

En la figura 4.6 se puede observar que a pesar de que se tienen algunas zonas que se confunden dentro de la matriz, se mejoró notoriamente el comportamiento anterior. Ya no es algo que podría decirse aleatorio sino que ahora está más cerca del comportamiento que se busca. Aquí se puede ver que se tienen algunas otras zonas que se confunden aparte de las mencionadas anteriormente, esto se debe a que teniendo menos APs se tiene menor precisión en la clasificación de las huellas. Para este caso se pudo alcanzar una precisión de 92,1 %. No es la misma precisión que se encontró con todos los APs pero sin dudas que es una gran mejora contra el resultado de clasificar con los clasificadores que tienen en cuenta todos los APs. También este valor supera el criterio de aceptación. Se puede ver entonces que la solución planteada es una posibilidad en cuanto a rendimiento de la misma. Una de las desventajas de esta solución sería en cuanto a lo económico, dado que tener la solución duplicada significa tener que incurrir en más gastos para la infraestructura. Sin embargo, con una reestructura del código de la aplicación existe la posibilidad de mitigar esto y crear una arquitectura que no tenga que duplicar la infraestructura.

Capítulo 4. Análisis de desempeño

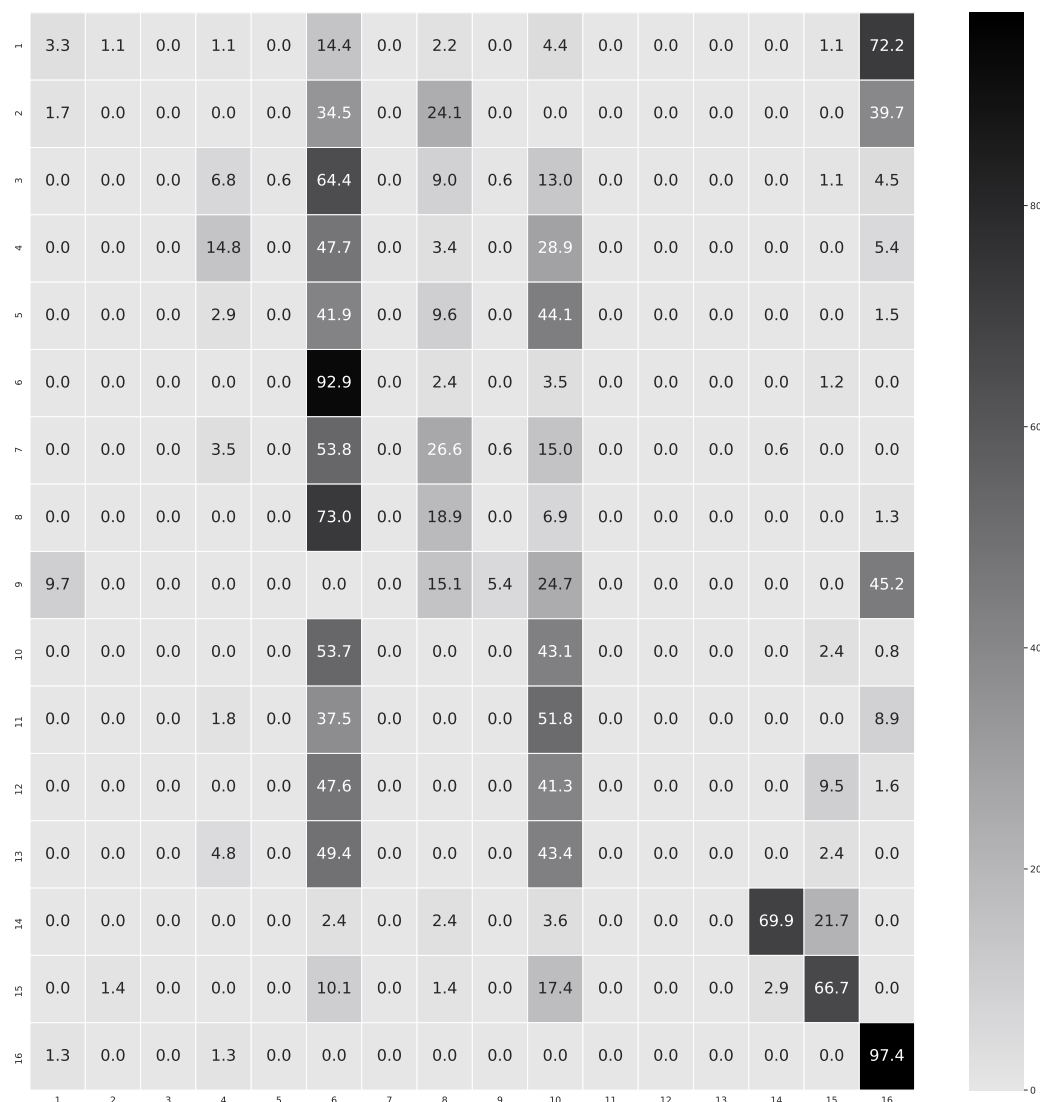


Figura 4.5: Matriz de confusión de entrenamiento en doble banda y clasificación en banda de 2.4GHz

4.4.3. Precisión en función de cantidad APs instalados

A partir del resultado con respecto a utilizar los APs conocidos (se concluyó que el uso de estos APs exclusivamente es beneficioso) se buscó obtener una relación entre la cantidad de APs utilizados para comprobar el sistema y la precisión. Este resultado es de gran importancia en cuanto a análisis de costos, teniendo en cuenta que la compra e instalación de los APs constituye quizá el mayor gasto del proyecto. Cada uno de los APs tiene un costo de unos cientos de dolares, por lo cual reducir la cantidad de los mismos ayuda a reducir los costos del proyecto considerablemente. Un análisis en este sentido puede proporcionar una idea de cuántos APs son realmente necesarios para lograr distintos niveles de precisión.

4.4. Entrenamiento con selección de MACs

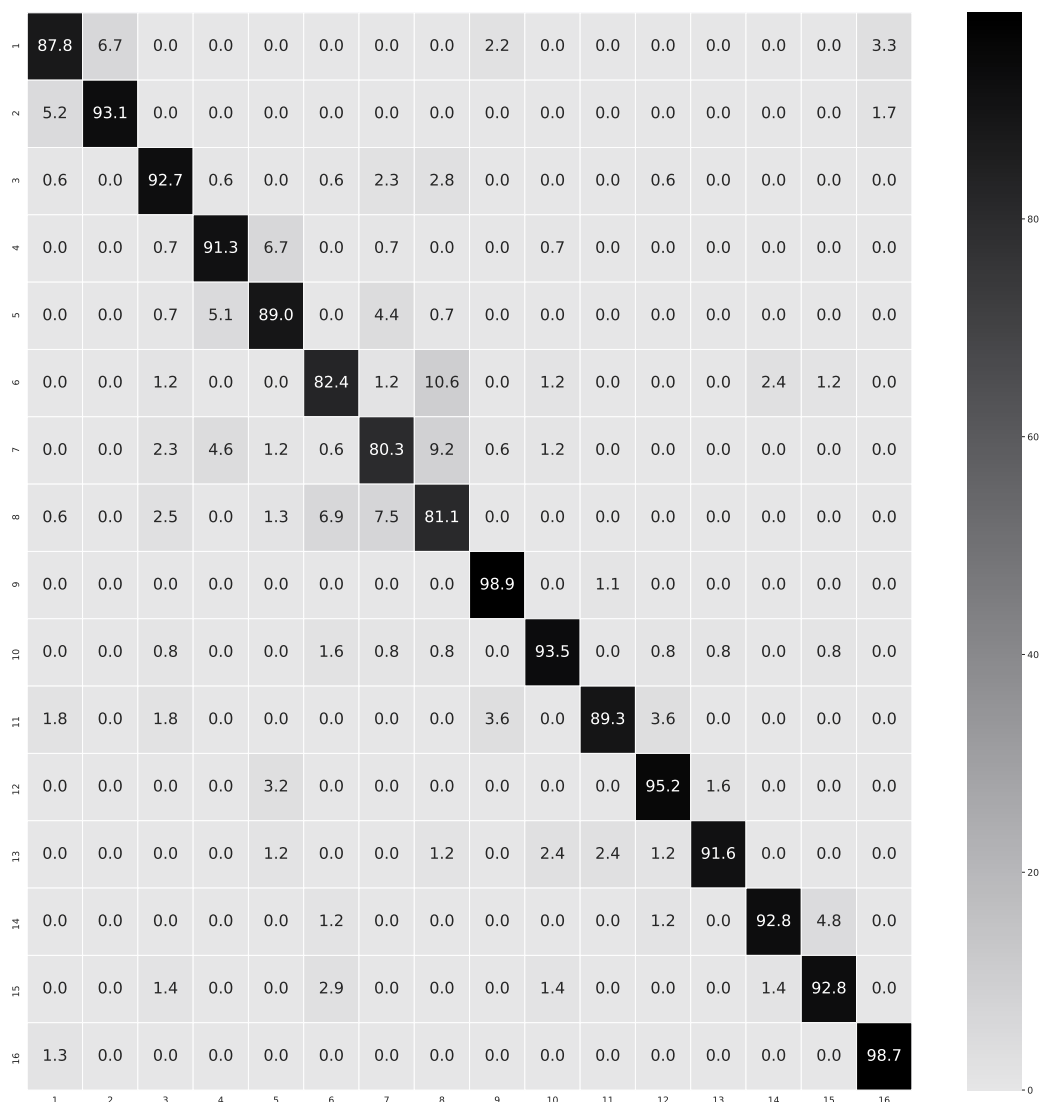


Figura 4.6: Matriz de confusión de entrenamiento y clasificación en banda de 2.4GHz

Para realizar este estudio, se filtraron todos los datos de forma que se fuera eliminando de a un AP, hasta eliminar 5 APs en total. Los APs a eliminar fueron elegidos bajo el criterio de eliminar los más redundantes en cobertura de señal primero. De esta forma, se puede de cierta forma simular el escenario en donde se planifique inicialmente la instalación con una menor cantidad de APs. Para decidir qué APs resultaban más redundantes, se basó en el mapa 3.6 acorde a como terminaron distribuidos. Se tomaron suposiciones al momento de elegir qué APs son los más redundantes, ya que no se realizó un análisis profundo del impacto de cada uno en la cobertura. Los APs fueron descartados en el siguiente orden:

- AP13 : Se encuentra en la sala de conferencias. Alejado de las zonas de localización.

Capítulo 4. Análisis de desempeño

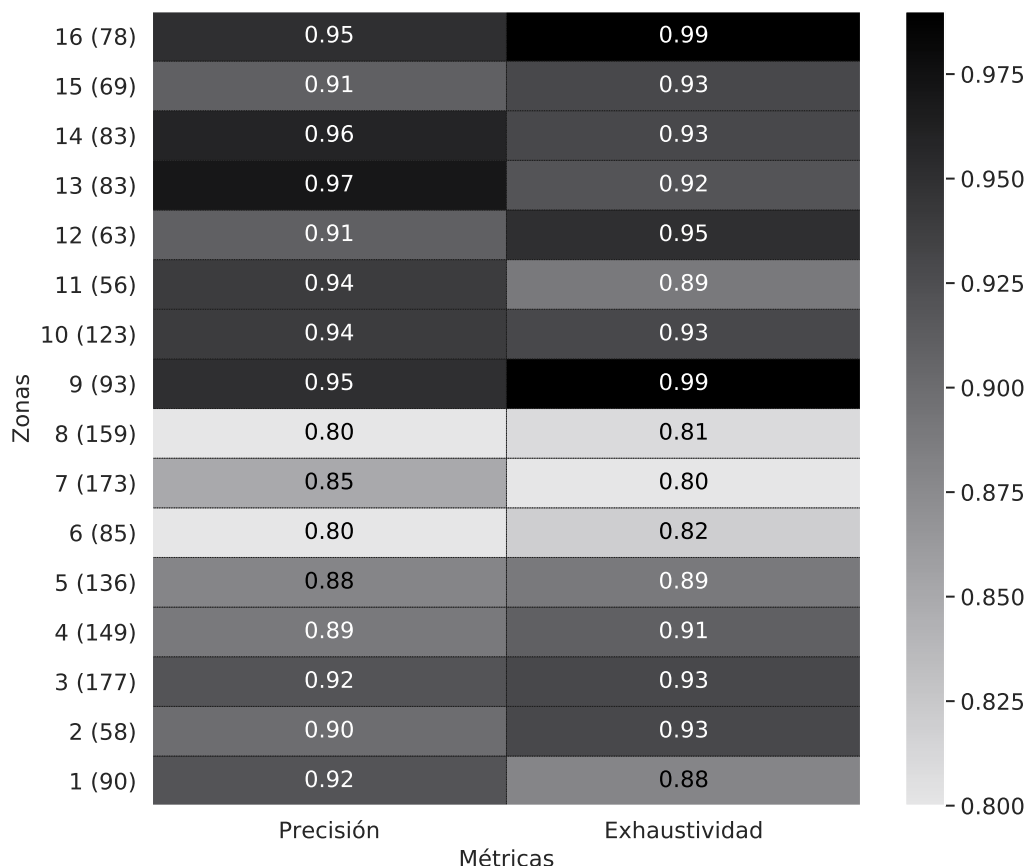


Figura 4.7: Tabla de cobertura y precisión de entrenamiento y clasificación en banda de 2.4GHz

- AP10 : Se encuentra en el depósito de obras, alejado de las zonas de localización.
- AP14 : Se encuentra en la sala de dirección, no directamente dentro de zonas de localización.
- AP3 : Se encuentra en una zona alejada con varios APs.
- AP8 : Se encuentra entre el AP7 el AP9, podría entenderse que es redundante a ellos dos.

En el caso extremo de solo contar con 10 APs el valor obtenido fue 94,48 %, un valor por encima del criterio de aceptación. En la figura 4.8 podemos ver los valores para las distintas cantidades de APs. A partir de esto podemos afirmar que se puede construir una solución robusta con menor cantidad de APs. Es preciso volver a señalar que la elección no fue arbitraria y se realizó a conciencia, teniendo en cuenta la redundancia de los APs. Por esto se puede afirmar que si se cuenta con menor cantidad de APs, haciendo una buena distribución y diseño de despliegue de los mismos se puede llegar a una solución que cumple con los valores de aceptación.

4.4. Entrenamiento con selección de MACs

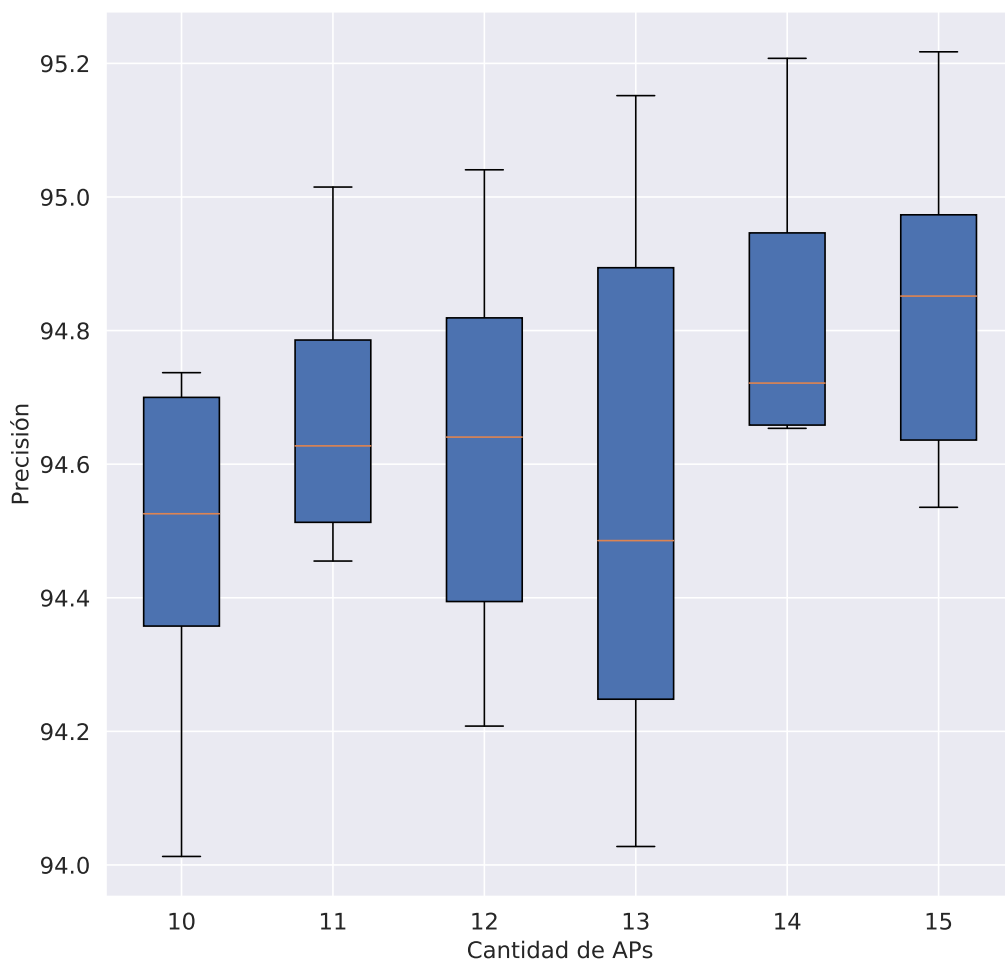


Figura 4.8: Precisión en función de cantidad de APs utilizados en la instalación. Se muestra en cada punto el boxplot con las diez simulaciones por conjunto de APs sacados.

De esta forma, es posible reducir los costos del sistema a expensas de obtener menos precisión pero obteniendo un sistema con rendimiento satisfactorio igualmente.

En búsqueda de llegar a un resultado extrapolable a futuros proyectos con distintas características, se buscó estudiar cuantos APs son visibles en cualquier punto del museo utilizando 10 APs. Esto permitiría tener una mejor idea de cuanta densidad de APs es necesaria independientemente del tamaño o estructura del ambiente. Luego de realizar el filtrado anterior dónde se eliminaron los APs de entrenamiento, se estudió el csv resultante, buscando aquellas huellas con menor cantidad de APs visibles. Se encontró una única huella con tan sólo 1 AP visible, sin embargo esta se descartó como anómala por representar menos que 0,01 % de los datos. Por otro lado, se encontraron varias huellas con 4 señales visibles, donde 3 de ellas correspondían a 2.4GHz y 1 a 5GHz. Esto permite pensar que para un nuevo proyecto se puede planificar de forma de que al menos 3 APs sean visibles en todo punto de la estructura.

4.4.4. Precisión en función de posibles fallas en APs

Mientras que el análisis anterior estudió el efecto de entrenar y utilizar el sistema con una menor cantidad de APs, otra posibilidad que vale la pena estudiar es el impacto que tiene eliminar un AP luego de entrenado el sistema simulando así una falla en dicho AP. En este estudio se plantea estudiar la disminución de precisión del sistema a medida que se van apagando hasta 3 APs cualesquiera, generando que la clasificación se lleve a cabo con menor información que la entrenada. Se eligió este número dado que la probabilidad de que se rompan más de 3 APs en una instalación es muy baja. Este análisis provee una imagen clara de la robustez del sistema frente a fallos en APs particulares.

Para poder observar estos cambios, se entrenó el sistema con los 15 APs instalados, para luego clasificar con los mismos datos de validación filtrados eliminando algunos APs. Para cada número de APs se promedió con todas las combinaciones posibles. Es decir, en el caso de eliminar 2 APs, se promedió la precisión final con los datos resultantes de eliminar cualquier pareja posible de APs.

A medida que se redujo la cantidad de APs fue disminuyendo el valor de precisión. El menor valor que se obtuvo con 3 APs faltantes fue de 93,7 %. Por lo cual se disminuyó aproximadamente un 2 % de la solución total. Sin embargo, este resultado sigue estando por encima del criterio de aceptación y dentro del intervalo error de estimación. Esto nos dice nuevamente que a pesar de tener pérdidas por fallas en los APs el sistema sigue siendo funcional y con un rendimiento apropiado.

4.5. Precisión en función de cantidad de muestras

El proceso de obtener las huellas constituye otro gasto importante en la forma de horas de trabajo. Para el caso del museo, la aplicación diseñada y utilizada para recolectar huellas obtiene una huella cada 2 segundos. Se obtuvieron 10469 huellas para el sistema final y se habían tomado alrededor de la misma cantidad en pruebas previas a la disposición final del museo, por lo que la cifra de huellas totales tomadas para la construcción del sistema es de alrededor de 20 mil huellas. Teniendo en cuenta los errores cometidos y corregidos (borrando y recalibrando zonas) y los tiempos entre tomar referencias y comenzar a medir se puede considerar que cada huella se toma en un tiempo efectivo de alrededor 2,2 segundos, 10 % por encima de obtener la huella solamente. De esta forma se tiene que el tiempo total en horas de recolección de datos fue:

$$Tiempo = \frac{22,000 \times 2}{3600} = 12 \text{ horas } 13 \text{ minutos.} \quad (4.1)$$

Este tiempo requiere tanto del personal necesario, el dispositivo y el acceso a las instalaciones que ya se deben encontrar en condiciones similares a las que va a regir el sistema.

De la misma forma que se estudió la cantidad de APs necesarios para obtener una cierta precisión del sistema, es importante entonces aplicar el mismo razonamiento al relevamiento de huellas.

4.5. Precisión en función de cantidad de muestras

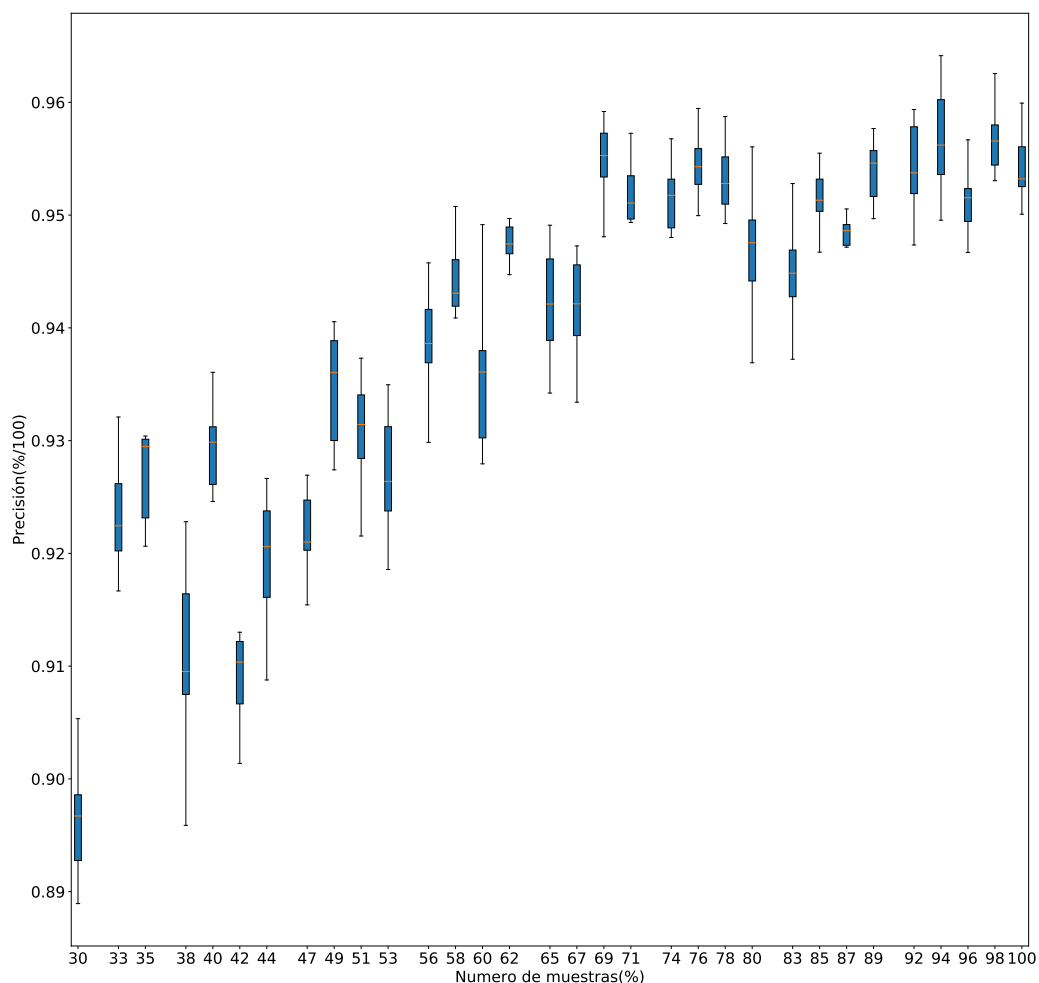


Figura 4.9: Precisión en función de cantidad de muestras. Se utiliza el boxplot para graficar las 10 simulaciones.

Para esto se calculó la precisión del sistema para distinta cantidad de huellas. Se varió entre 30 % y el 100 % de las muestras. Para obtener las fracciones de las huellas se siguieron algunas reglas.

- Se mantiene la proporción de cantidad de huellas en cada zona por sobre la cantidad de huellas total. Las zonas mantienen la misma relación de huellas entre ellas.
- Las huellas a elegir para cada prueba se eligen aleatoriamente.
- Debido al carácter aleatorio que afecta la elección de los subconjuntos de huellas, la precisión se promedia sobre 10 elecciones.

En la figura 4.9 se puede observar la tendencia a bajar la precisión a medida que hay menor cantidad de huellas. De la misma forma que para el análisis de

Capítulo 4. Análisis de desempeño

cantidad de APs, se debe decidir la cantidad de huellas a tomar en cuanto al costo real de obtenerlas y la precisión final objetivo. Para asegurar alrededor de un 90 % de precisión se necesitarían por ejemplo alrededor de 2500 huellas. Es importante notar que la variabilidad de la elección de las huellas utilizadas y de los conjuntos de prueba dentro de ellas resulta en la forma irregular de la gráfica.

4.6. Resolución espacial

En la sección 3.5 se describieron los distintos criterios que se tomaron para definir las zonas. En algunos sentidos, estos se ajustaban a la aplicación específica y hacían uso de elementos característicos de la instalación del museo. Sin embargo, estos elementos se realizaron sin un análisis concreto de la capacidad de resolución del sistema, por lo cuál los tamaños de las zonas se tomaron de un área mayor para ser conservador. En esta sección se plantea realizar un estudio de cuál es la resolución alcanzable realizando un entrenamiento con zonas de diverso tamaño en forma de cuadrícula. De esta forma, se podría dar una idea de cuánto puede sacrificar la precisión del sistema con tal de mejorar la resolución del mismo.

Para realizar este estudio no se pudo utilizar los datos recabados anteriormente para la muestra, ya que en estos ya se definieron las zonas. Esto condujo a que se levantara un sistema paralelo con la misma arquitectura pero con nuevas huellas. Se trabajó únicamente en planta baja, lo cuál eliminó las confusiones en cuanto a zonas de distintos pisos, pero simplificó el planteo de este estudio concreto.

La distribución de zonas se realizó como se observa en la figura 4.10. Estas zonas que se llamarán “mínimas” crean el punto de partida. Utilizando etiquetas de la forma l_{xy} , se obtienen huellas para estas zonas, donde x es la coordenada horizontal y y es la coordenada vertical. Para analizar el efecto de utilizar zonas más grandes, se decide llevar a cabo un proceso de unión de zonas adyacentes. Se iterará en la unión de las zonas hasta tener un valor parecido a las zonas desplegadas en el museo. En total se tomaron 7971 huellas, con zonas de entre 43 y 114 huellas dependiendo el tamaño y accesibilidad de las mismas.

Se tomaron dos enfoques distintos al problema de resolución espacial, el primero de ellos intentó hacer una estimación sin tener en cuenta la forma como se agrupaban las zonas. El segundo análisis fue hecho tomando en cuenta la elección de zonas previamente realizado.

Para hacer la elección de las nuevas zonas en el primer caso se unieron zonas adyacentes. Para el caso base se agruparon dos zonas mínimas. Luego se fueron juntando las zonas resultantes en grupos de a 2 hasta obtener 13 zonas que es el mínimo que permite la agrupación de esta forma. Las agrupaciones se realizaron tanto horizontalmente como verticalmente en dos casos separados para probar las diferencias entre ambas. Para el segundo caso, se dividieron las zonas previamente definidas de forma que la subdivisión de las zonas tenga sentido con la división original de las zonas.

Podemos observar en la figura 4.11 que a pesar de haberse acercado a el numero de zonas que se tiene el el sistema verdadero, esto no alcanzó para obtener una buena precisión.

4.6. Resolución espacial

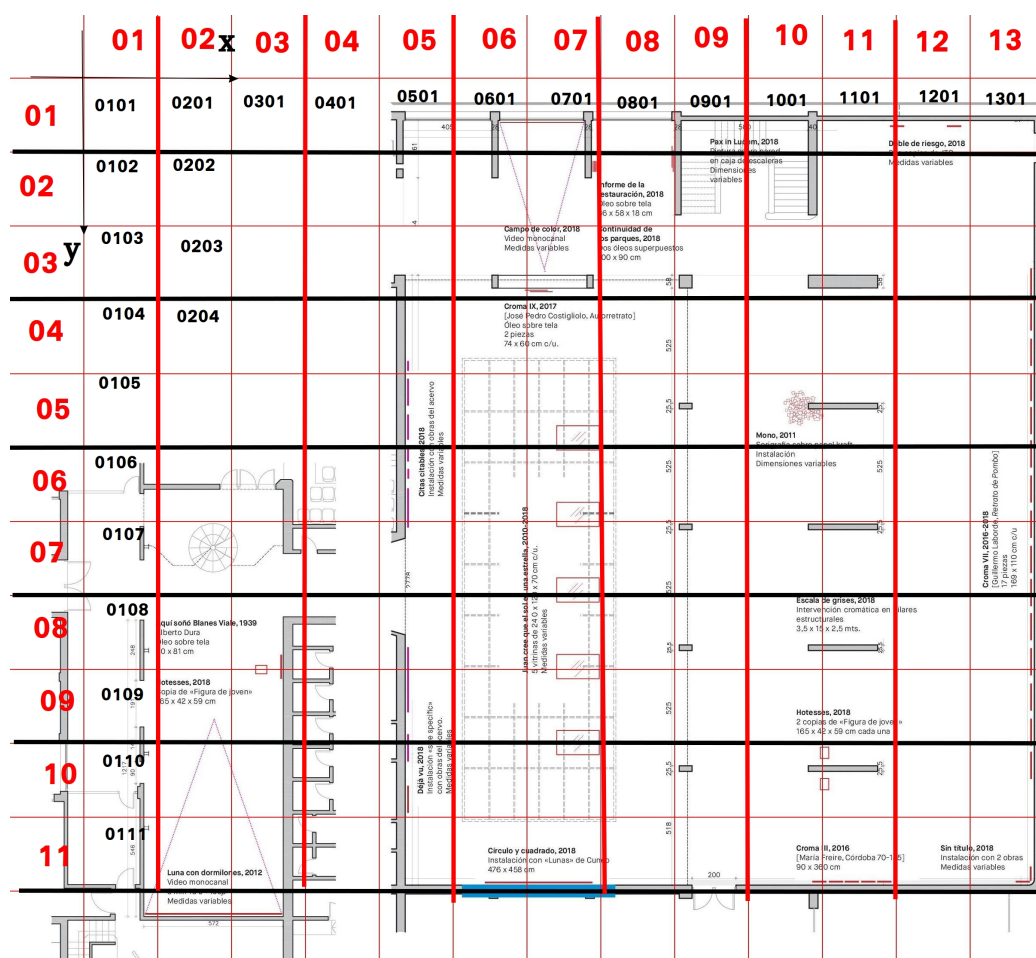


Figura 4.10: Distribución de zonas mínimas para estudio de resolución espacial.

Se observa que la precisión aumenta significativamente a medida que disminuimos el número de zonas, llegando a una diferencia mínima entre las dos últimas divisiones. Más allá de ello y teniendo en cuenta que se tomaron 7971 huellas, se nota que la precisión es mucho menor a la del sistema en producción que ronda 95%. Hay dos factores principales que se creen pueden haber sido determinantes en esta diferencia.

El primer factor se puede deber a que en el relevamiento de estas zonas de menor tamaño, se obtuvieron una pequeña cantidad de huellas por zona.

Por otro lado, la distribución de las zonas realizada en una cuadrícula no teniendo puntos de referencia presentó una mayor dificultad al momento de relevar las huellas. Teniendo en cuenta que el plano no coincide muy precisamente con el museo, el relevamiento de huellas presentó mayor cantidad de errores humanos. Siendo el principal de ellos el relevamiento de huellas con etiquetas equivocadas.

Más allá de los posibles errores, el tiempo requerido para tomar estas huellas aumentó de gran manera. El tiempo estimado para obtener una huella en el caso del sistema en producción teniendo en cuenta el ajuste de la aplicación y el repo-

Capítulo 4. Análisis de desempeño

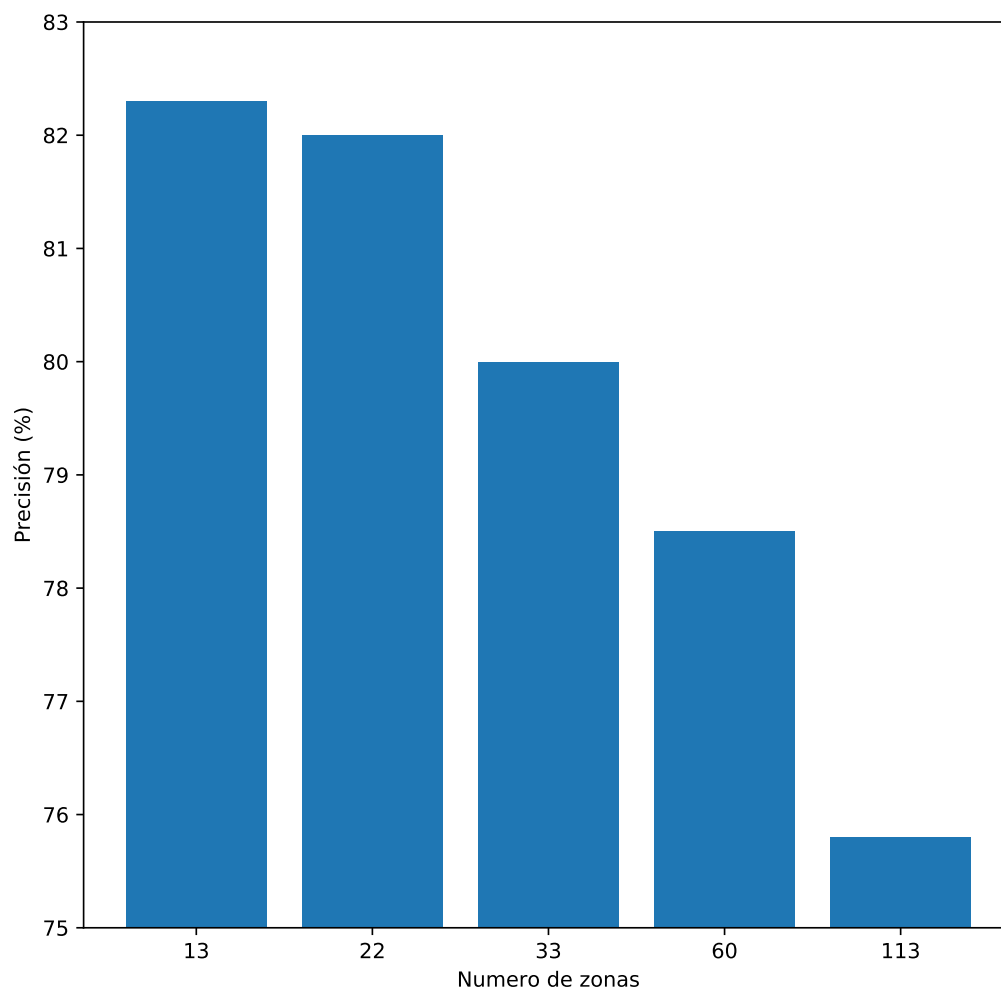


Figura 4.11: Precisión general obtenida según la cantidad de zonas generadas en la distribución del espacio de planta baja.

sicionamiento se considera de 2.2 segundos (un factor de 1.1 con respecto a los 2 segundos por huella de la aplicación). En el caso de las huellas para la experiencia de resolución especial, cada huella tomó efectivamente alrededor de 3.2 segundos (alrededor de un factor de 1.6). Esto eleva el costo de horas hombre necesarias en esta labor.

Otro factor a considerarse es la aplicación. Dada la experiencia y los comentarios recibidos por los usuarios, se observó que a veces se encontraban insatisfechos con no poder visualizar cierta obra a una distancia considerable. Esto es debido a que a pesar de tomar zonas grandes, las obras suelen ser vistas desde muy diversas posiciones. Es importante entonces en esta aplicación recordar que una mayor resolución espacial no mejoraría la experiencia de usuario final.

Por el contrario de este resultado, cuando se realizó la experiencia para la elección de zonas no arbitrarias, se pudo alcanzar una precisión del 89 % lo cual esta

4.7. Resumen de resultados

mucho más cerca del valor del sistema real. Otra vez como se planteó anteriormente podemos concluir a partir de este resultado que los errores humanos afectaron la calidad de este dataset, disminuyendo el valor de la precisión en 5 puntos porcentuales. A pesar de esto cualitativamente vemos que haber elegido inteligentemente las zonas mejoró la precisión sustancialmente.

Teniendo todo esto en cuenta se observó que intentar tener mayor resolución en el sistema a expensas de reducir las zonas no es un camino que se pueda realizar para cumplir con el criterio de aceptación cuando se hace arbitrariamente. Podemos concluir de este análisis que la elección no arbitraria de las zonas es un factor importante a la hora de decidir la distribución de las zonas. Teniendo una elección adecuada y a conciencia de las mismas da una mejor resultado que realizar la división de las zonas arbitrariamente. Esto valida entonces que la primera elección de las zonas fue una buena decisión de diseño.

4.7. Resumen de resultados

Para una mejor visualización de los resultados obtenidos en el análisis de datos del sistema, se presentan las precisiones obtenidas durante la distintas pruebas.

- Precisión para la totalidad de los datos
 - Muestras con datos faltantes impuestos en $0dBm$: 95,7 %
 - Muestras con datos faltantes impuestos en $-90dBm$: 96,8 %
- Precisión para datos solo con APs conocidos: 95,6 %
- Precisión para datos entrenados con huellas solo en la banda de 2 GHz: 92,1 %
- Precisión para entrenamiento con todas las huellas y clasificación con celulares que tienen presente solo una banda: 30 %
- Precisión para entrenamiento con todas las huellas y clasificación habiendo faltante de APs
 - Para 1 AP faltante: 94,3 %
 - Para 2 APs faltantes: 94,0 %
 - Para 3 APs faltantes 93,7 %
- Precisión para entrenamiento con APs faltantes: Usando 10 de los APs se logra 93,5 %

Esta página ha sido intencionalmente dejada en blanco.

Capítulo 5

Estadísticas de uso

5.1. Introducción

Una de las motivaciones principales de los sistemas de localización indoor siempre ha sido la posibilidad de generar estadísticas respecto a los recorridos de los usuarios del sistema. Este tipo de información se puede presentar cuantitativamente en función de horarios de mayor actividad así como zonas más concurridas.

En el caso del proyecto del MNAV llevado a cabo para la muestra Aquí soñó Blanes Viale, se recolectaron los datos durante los tres meses de la misma.

Las respuestas del sistema a los usuarios que utilizan MNAVegante cuentan con un campo que identifica al usuario según su dispositivo, un campo con la zona en la que se encontraba y un campo de marca de tiempo el cuál permite ubicarlo en el tiempo.

Otra información que es brindada por la plataforma de Google Play, donde fue publicada la aplicación, son las estadísticas de uso de la misma, así como su rendimiento.

Este es un factor importante para poder analizar el uso y adquisición de la misma. Con estos datos se pudo estimar la cantidad de usuarios a los que se llegó con la aplicación y se pudieron detectar en tiempo real las fallas de la misma.

5.2. Actividad de usuarios en el tiempo

Utilizando el campo de tiempo, se puede realizar un conteo básico de cuántas solicitudes de localización se llevaron a cabo por día de la semana y por hora del día.

En la figura 5.1 se puede observar enseguida que el lunes tiene una mínima cantidad de actividad. Esto se debe a que el museo permanece cerrado los lunes. Por otro lado los jueves parecen ser los días de mayor tráfico con más del doble de datos que el siguiente día. Esto sin embargo fue influenciado por el hecho de que la apertura de la muestra (evento de gran tráfico en la app) sucedió un jueves obteniéndose una mayor concurrencia en ese día. Esto se pudo comprobar removiendo los datos correspondientes a la fecha de la apertura de la muestra. En la

Capítulo 5. Estadísticas de uso

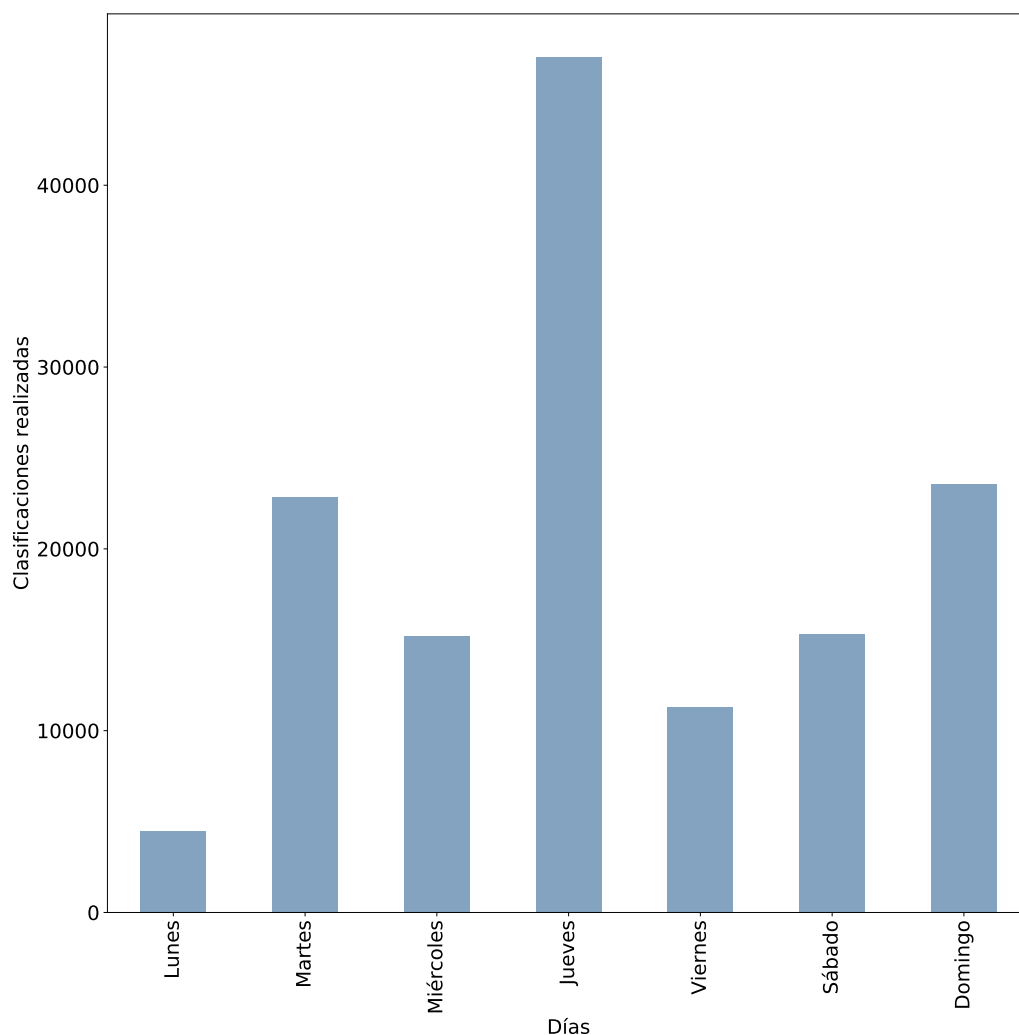


Figura 5.1: Clasificaciones realizadas por día de la semana

figura 5.2 se aprecia el efecto que tiene remover estos datos, ya que el jueves pasa a ser el tercer día con mayor actividad. Por otro lado, se destacan los domingos y martes como los días con mayor uso de la aplicación.

Por otro lado, en la figura 5.3 se puede claramente discernir los horarios de apertura y clausura del museo, abriendo este a las 13 horas y cerrando a las 20 horas. Se destaca el horario entre las 18hs y las 19hs como el período de mayor tráfico. El tráfico existente a las 21hs en adelante, corresponde a las instancias como la apertura de la muestra o la fecha de museos en la noche cuando el museo permaneció abierto pasadas las 20hs.

Entonces podemos decir que los datos que obtuvimos de las estadísticas de los usuarios se condicen con la realidad del museo, por lo cuál este análisis es extrapolable para cualquier otra ubicación donde el sistema esta presente. Esto brinda una capa más de información que luego puede ser beneficioso para tomar

5.3. Zonas más concurridas

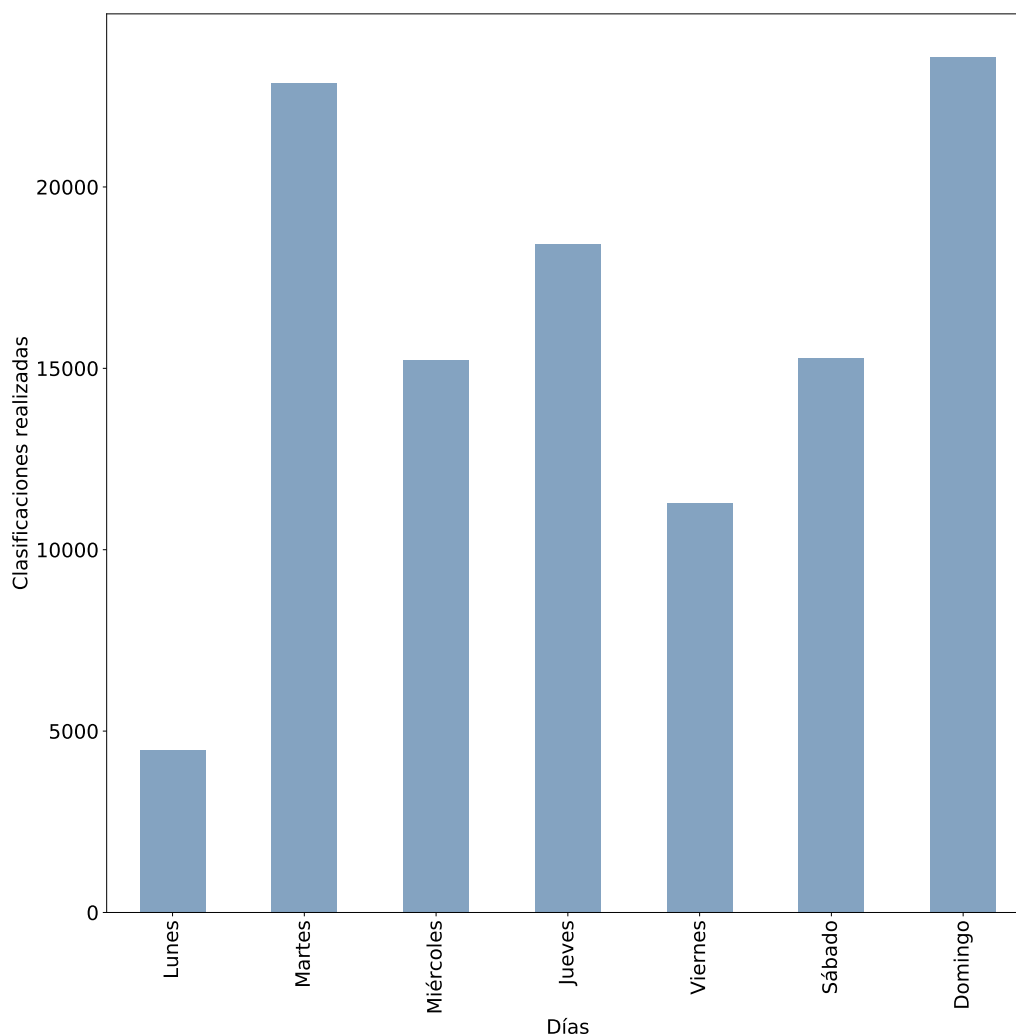


Figura 5.2: Clasificaciones realizadas por día de la semana quitando los datos obtenidos el día de la apertura.

decisiones en cuanto a los usuarios.

5.3. Zonas más concurridas

Utilizando la información de zona en los datos, se puede generar un conteo de en qué zonas se encontraban los usuarios mientras usaban la app. De esta forma, se puede rápidamente obtener una estadística de zonas más concurridas.

En la figura 5.4 se pueden observar las zonas más concurridas en función de cuántas clasificaciones resultaron en cada zona durante la muestra. La zona 10 es la más concurrida, seguida de la zona 6. Se pudo concluir que esta estadística está fuertemente sesgada por los dispositivos que no cuentan con banda de 5GHz debido a su mala clasificación como se mencionó en la parte de análisis de datos.

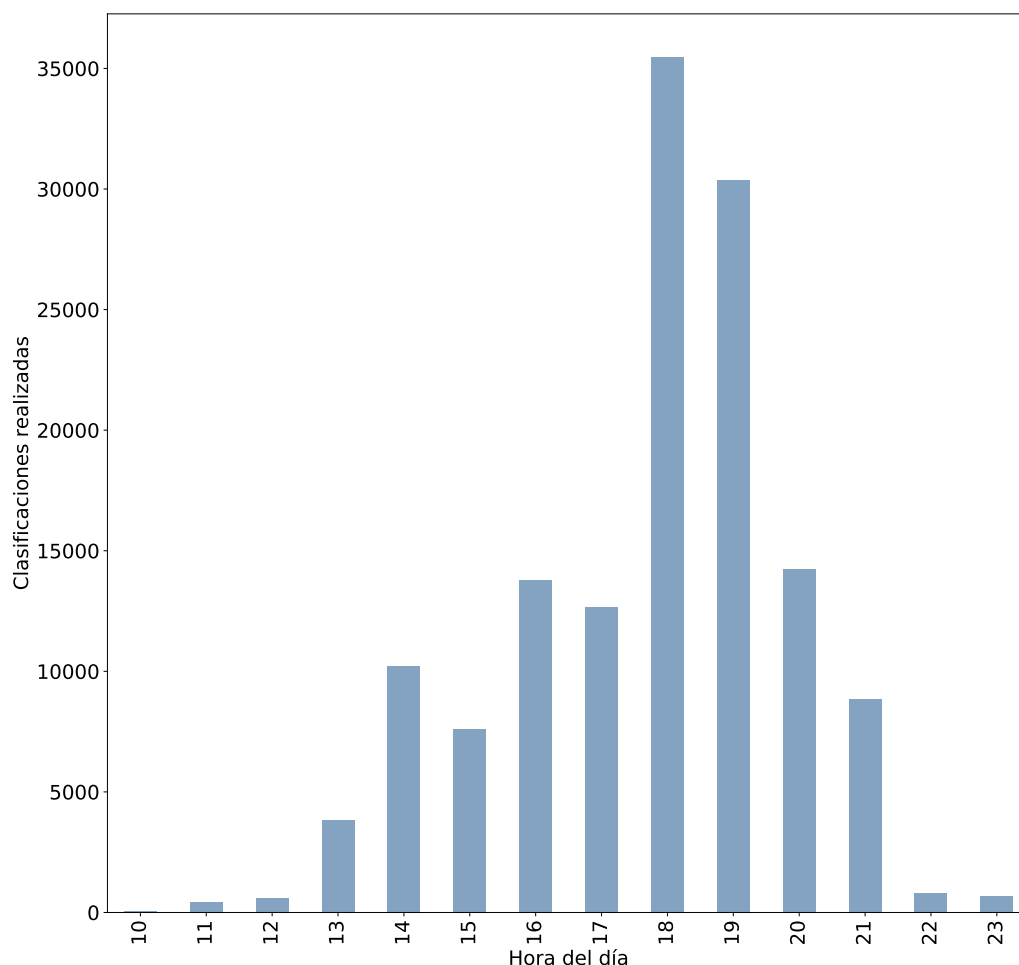


Figura 5.3: Clasificaciones realizadas por hora del día

Tal como se observó previamente en la matriz de confusión de la figura 4.5, los dispositivos que no cuentan con banda de 5GHz suelen generar predicciones en las zonas 6, 10 y 16.

Obviando las zonas más sesgadas por estos dispositivos, se puede observar que las zonas con mayor concurrencia fueron las zonas 4, 5 y 7 que componen gran parte de la planta baja, zona donde generalmente confluye la mayor cantidad de gente por tener las obras principales del museo.

5.4. Estadísticas de la aplicación

Utilizando la información proporcionada por Google Play se pueden sacar datos de como se comportó la aplicación en los meses que estuvo activa.

En la figura 5.5 se pueden observar la cantidad de descargas de la aplicación en el tiempo. Hay un pico en noviembre ya que es cuando se lanza la muestra Aquí soñó Blanes Viale, y es también cuando más difusión se le hizo a la misma.

5.4. Estadísticas de la aplicación

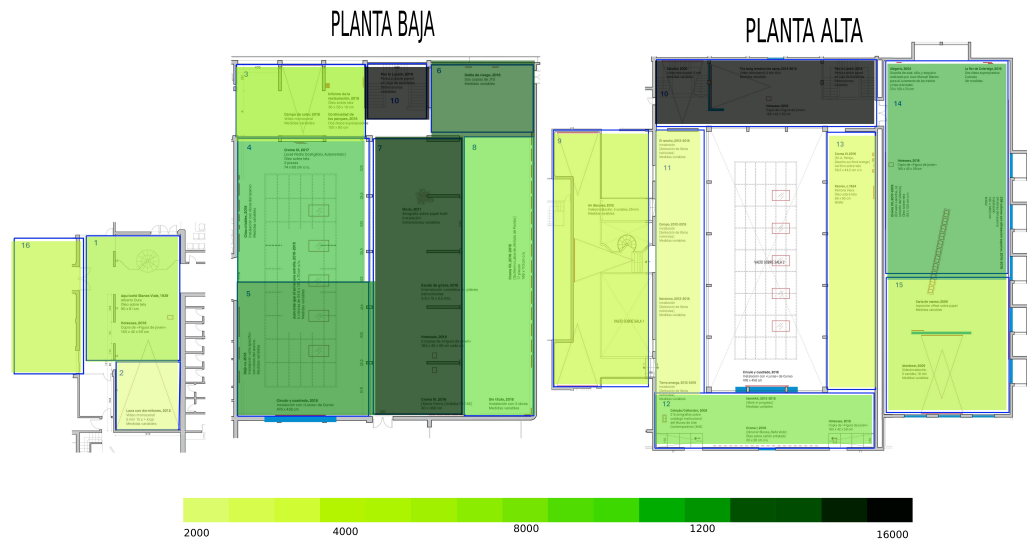


Figura 5.4: Mapa con conteo de personas por zona.

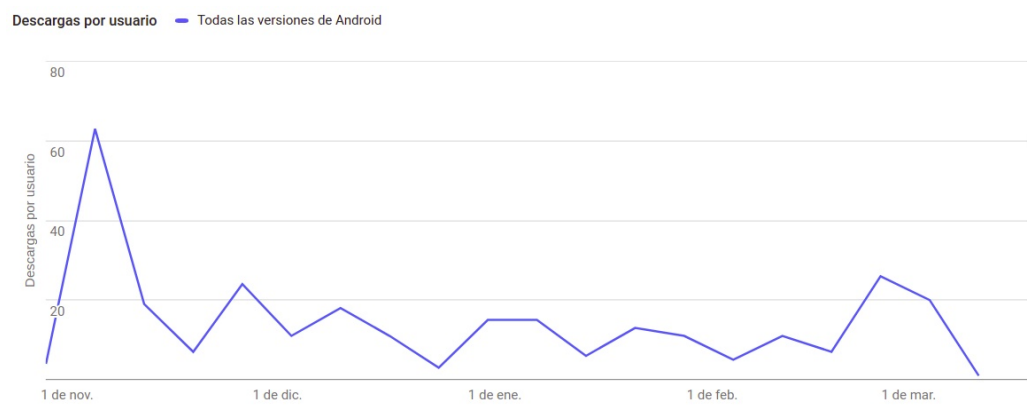


Figura 5.5: Cantidad de descargas por usuarios únicos en el tiempo.

Capítulo 5. Estadísticas de uso

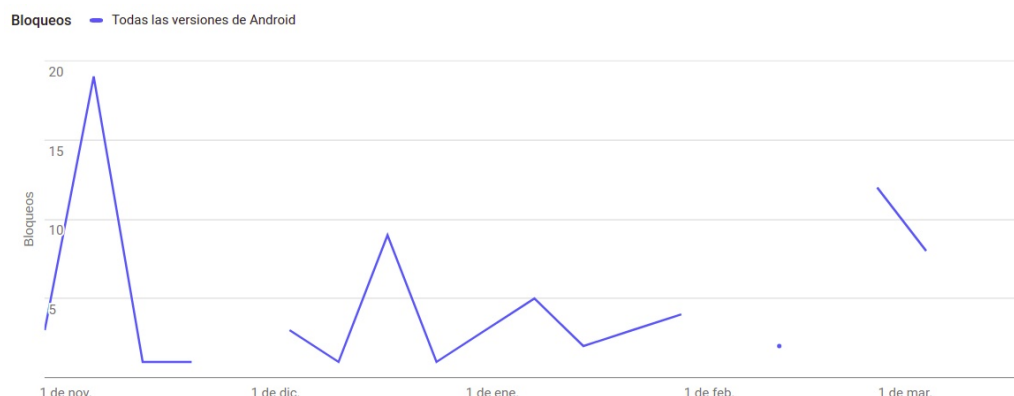


Figura 5.6: Bloqueos en el tiempo a partir de estadísticas de Google Play.

Se puede observar otro pequeño pico a principio de Marzo, ya que fue cuando la muestra cerró y de nuevo, tuvo propaganda y fue bastante concurrida. Este análisis estilo permitió que luego de finalizado la muestra se pudiera observar el alcance que se tuvo.

En la figura 5.6 se pueden observar la cantidad de bloqueos sobre el tiempo. Un bloqueo es una salida abrupta de la aplicación, lo que puede traducirse como un error. Se puede apreciar que en noviembre se concentra la mayor cantidad de errores de la aplicación. Dicha fecha coincide con el lanzamiento de la aplicación, esto es razonable, ya que fue la primera vez que la misma fue expuesta al público. En esta fecha se detectaron los errores descritos en la sección 3.7, siendo que los bloqueos por pérdida de conectividad se ven reflejados en estas estadísticas.

Capítulo 6

Conclusiones y líneas de trabajo a futuro

6.1. Conclusiones

6.1.1. Introducción

En el presente proyecto se logró el desarrollo y la implementación de un sistema de localización indoor basado en Wi-Fi, a partir de la utilización del framework libre FIND3. Este sistema se vio acompañado por una aplicación desarrollada para Android donde se hacía uso de la capacidad de localización para proveer un servicio a los visitantes de la muestra Aquí soñó Blanes Viale en el Museo Nacional de Artes Visuales.

6.1.2. Experiencia de usuario y recepción del público

El sistema y la aplicación probaron ser exitosos, logrando un buen desempeño en el contexto de proporcionar a los usuarios una experiencia interactiva en sus recorridos por la muestra. Esto permite la primera conclusión relevante, en cuanto a que el uso de la tecnología de localización basada en Wi-Fi es una posibilidad factible para este tipo de aplicaciones.

La experiencia de los usuarios y la recepción por el público en general fueron positivas y destacaron el aspecto novedoso de la tecnología utilizada. A modo de ejemplo una reseña realizada por un usuario en la Play Store destaca “Excelente propuesta. Moderna, dinámica y accesible para todo público”. Este no solo refleja un correcto funcionamiento sino las mayores virtudes del proyecto en cuanto a haber logrado sus planteamientos en el área de experiencia de usuario y accesibilidad.

La aplicación MNAVegante y la muestra Aquí soñó Blanes Viale contaron con difusión en los medios, tanto en diarios como en blogs. Se destaca por ejemplo un artículo realizado por el Ministerio de Diseño [5]. Un fragmento en particular permite destacar a la aplicación y la muestra en el conjunto que forman.

“En esta experiencia en la que parte del acervo del arte nacional conversa den-

Capítulo 6. Conclusiones y líneas de trabajo a futuro

tro de un mismo escenario tecnológico, la ingeniería da cuenta de un resquebrajamiento de la antigua relación contractual entre el museo, la obra y el espectador, para abrir las puertas a nuevas formas de contemplar, entender y vincularse. [5]”

A partir de la difusión lograda y el éxito de la muestra, el grupo de proyecto ha sido contactado por diversas partes interesadas en implementar un sistema similar. El interés generado refleja además la capacidad de esta tecnología de ser aplicada en diversos ámbitos, habiendo sido por ejemplo una de las partes interesadas un cementerio.

6.1.3. Análisis de desempeño

Para evaluar el rendimiento del sistema tanto en la implementación real como en distintos escenarios hipotéticos se utilizaron los datos recabados durante la recolección de huellas del sistema, así como algunas huellas adicionales para nuevos análisis. Estos estudios proporcionaron una imagen más clara de cuales son los principales factores a tomar en cuenta para lograr un buen sistema de localización. Por un lado incluye el estudio de factores y variables que pueden afectar negativamente la precisión del sistema una vez instalado, tales como el efecto de APs que sean apagados o el filtrado de los datos utilizando tan solo los APs instalados para el proyecto. Uno de los mayores inconvenientes del sistema instalado fue el pésimo rendimiento de los dispositivos sin la banda de 5GHz. Este fenómeno fue estudiado en detenimiento, proponiendo la alternativa de utilizar un sistema paralelo entrenado con un dispositivo sin 5GHz.

Por otro lado se estudiaron distintas variables a nivel de planificación inicial. Estas incluyen la cantidad de huellas o cantidad de APs necesarios para lograr una cierta precisión objetivo. Estos dos estudios permiten estimar y planificar costos, debido a que la compra e instalación de APs y la recolección de huellas constituyen los mayores costos del proyecto. Por último, dentro de los estudios de planificación inicial, se consiguieron resultados respecto a la resolución espacial alcanzable bajo las mismas condiciones para tener un parámetro de como se debe plantear la distribución de zonas de localización en el ambiente.

Adicionalmente, se pudieron recolectar datos de los usuarios en forma de zonas más concurridas y horarios o días de la semana con mayor cantidad de visitantes. Más allá de verse sesgado el estudio por condiciones como los dispositivos que no cuentan con la banda de 5GHz, este tipo de información es uno de los aspectos que puede generar un valor adicional al espacio que utiliza el sistema como lo es en este caso el MNAV.

Las tecnologías de localización indoor se encuentran en pleno desarrollo a tal punto que la información disponible sobre el tema cambió rápidamente a medida que se desarrollaba este proyecto. El uso de estas tecnologías en edificios públicos como museos para proveer servicios no es una idea nueva. Sin embargo, la efectividad demostrada en la aplicación del proyecto demuestra que estas tecnologías ya tienen la madurez necesaria para un despliegue masivo y se espera que sean integradas como uso cotidiano en los próximos años.

6.2. Trabajo a futuro

6.2.1. Introducción

Esta sección pretende dar un panorama de cuáles son las posibles líneas de trabajo a futuro que se abrieron como posibilidades en el correr del proyecto. Entre ellas, se encuentran aquellos casos donde la experiencia obtenida permite repensar metodologías utilizadas para la implementación para el caso que se comenzara el proyecto de cero.

6.2.2. Dimensionamiento en cantidad de APs y huellas

En las secciones 4.5 y 4.4.3, se estudiaron los efectos sobre el sistema al utilizar distintas cantidades de APs o huellas. Estos dos aspectos conformaron los principales costos de despliegue, por lo cual en caso de una futura implementación se deben reducir lo más posible sin prescindir de demasiada precisión. Este estudio depende de las instalaciones utilizadas ya que está altamente ligado a las áreas a dar cobertura, a la resolución espacial requerida y a la estructura general del ambiente. Sin embargo, a partir de los criterios de aceptación de la sección 4.2.1 y los resultados de las secciones 4.5 y 4.4.3, se pudieron haber utilizado menos APs y menos huellas. Particularmente, la reducción a 10 APs presenta una precisión más que aceptable y la recolección de muestras más allá del 70 % no parece tener un gran impacto. Con esto en mente, en un nuevo proyecto de las mismas características se cree que basta con utilizar alrededor de 10 APs y 7000 huellas en total para obtener una buena precisión (encima de 90 %). Una generalización cualitativa realizada en esta misma sección permite expandir la planificación en cantidad de APs a otras instalaciones. Planificando una densidad de APs tal que 3 APs sean visibles como mínimo para cualquier punto de la instalación, se deberían lograr condiciones similares a las del museo con 10 APs. De esta forma se puede lograr una buena precisión para otros espacios diferentes al museo imponiendo solo 3 APs visibles en todo el ambiente, en vez de 5 APs como en el diseño inicial.

6.2.3. Solución a dispositivos sin doble banda

Tal como fue estudiado en el capítulo de análisis de datos, el pobre desempeño de los dispositivos sin banda de 5GHz se debe al entrenamiento del sistema con dispositivos de doble banda. Este efecto puede ser mitigado realizando un sistema paralelo donde se realice el entrenamiento utilizando solo medidas en la banda de 2.4GHz. Esto permitiría lograr una precisión aceptable (o al menos, tolerable) para los dispositivos sin 5GHz. Para evitar prescindir de la alta precisión con la que cuentan los dispositivos de doble banda, se mantendrían ambos sistemas en paralelo, redirigiendo aquellos sin 5GHz al nuevo sistema.

Capítulo 6. Conclusiones y líneas de trabajo a futuro

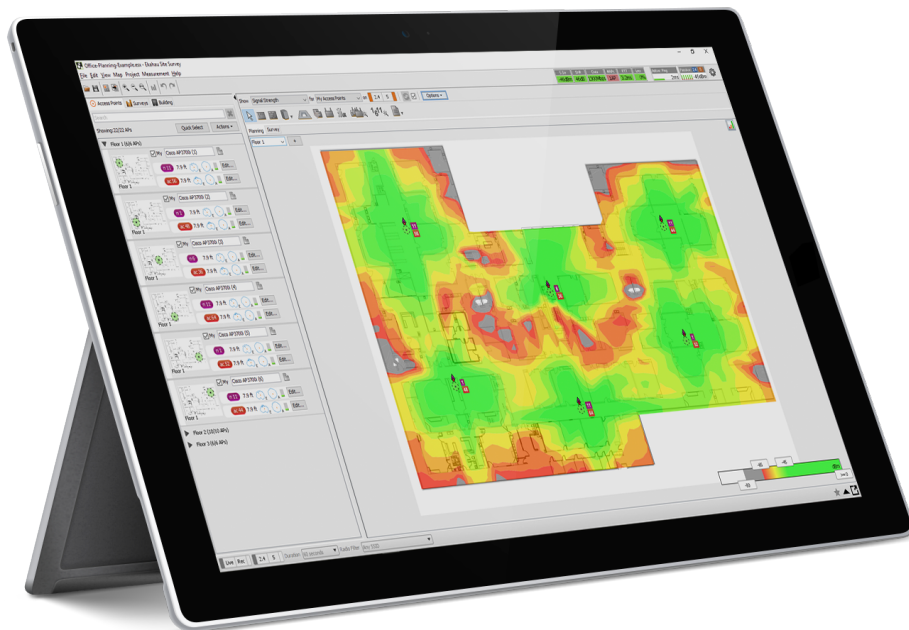
6.2.4. Filtro de Kalman

Una alternativa que fue planteada para hacer uso de la variable temporal consistiría en utilizar un filtro de Kalman. Esta información podría mejorar el sistema haciendo uso de la variable temporal, ya que un usuario se espera que se traslade de una zona a otra adyacente. A pesar de las ventajas que este sistema pudiera poseer, su desarrollo se vio imposibilitado por las limitantes de tiempo en cuanto a la apertura de la muestra. Luego de darse por comenzada la muestra, el análisis de datos y de rendimiento obtuvieron una mayor prioridad por lo cual su implementación no se realizó. Sin embargo, abre una línea de posibles trabajos a futuro.

Apéndice A

A.1 Reporte Ekahau

Wi-Fi Network Report

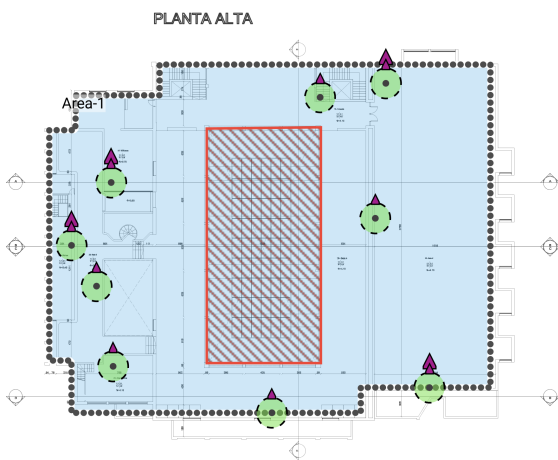


Network capacity configuration

	2.4 GHz	5 GHz
Minimum Data Rate	12 Mbits/s	12 Mbits/s
Band steering	N/A	N/A
Number of SSIDs	2	0
Max. Associated Clients	200	200
RTS / CTS	No	No

Planta_alta

Survey routes and Access Points for Planta_alta



Area-1 (104.6 m²)

Coverage Requirement: Posifi	Signal Strength Min	-80.0 dBm
	Signal-to-noise Ratio Min	8.0 dB
	Data rate Min	2 Mbps
	Number of Access Points Min	5 at min. -80.0 dBm
	Channel Overlap Max	3 at min. -80.0 dBm
	Round Trip Time (RTT) Max	500ms
	Packet Loss Max	10.0 %

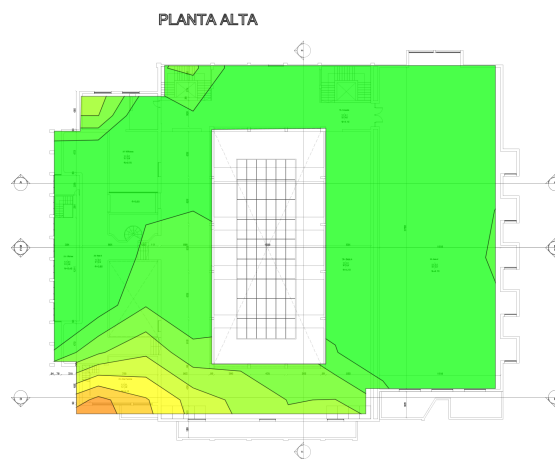
Wi-Fi Network Report

Capacity Requirement	100Generic Smartphone [Background Sync] Total: 100 (50 Mbits/s)
Notes	

Wi-Fi Network Report

Signal Strength for Planta_alta on 2.4 GHz band

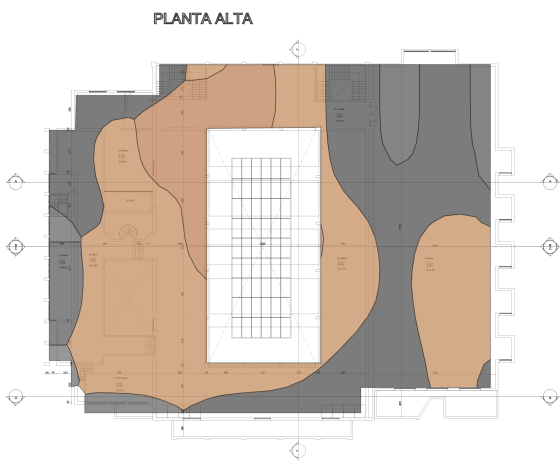
Signal Strength - sometimes called coverage - is the most basic requirement for a wireless network. As a general guideline, low signal strength means unreliable connections, and low data throughput.



Wi-Fi Network Report

Number of APs for Planta_alta on 2.4 GHz band

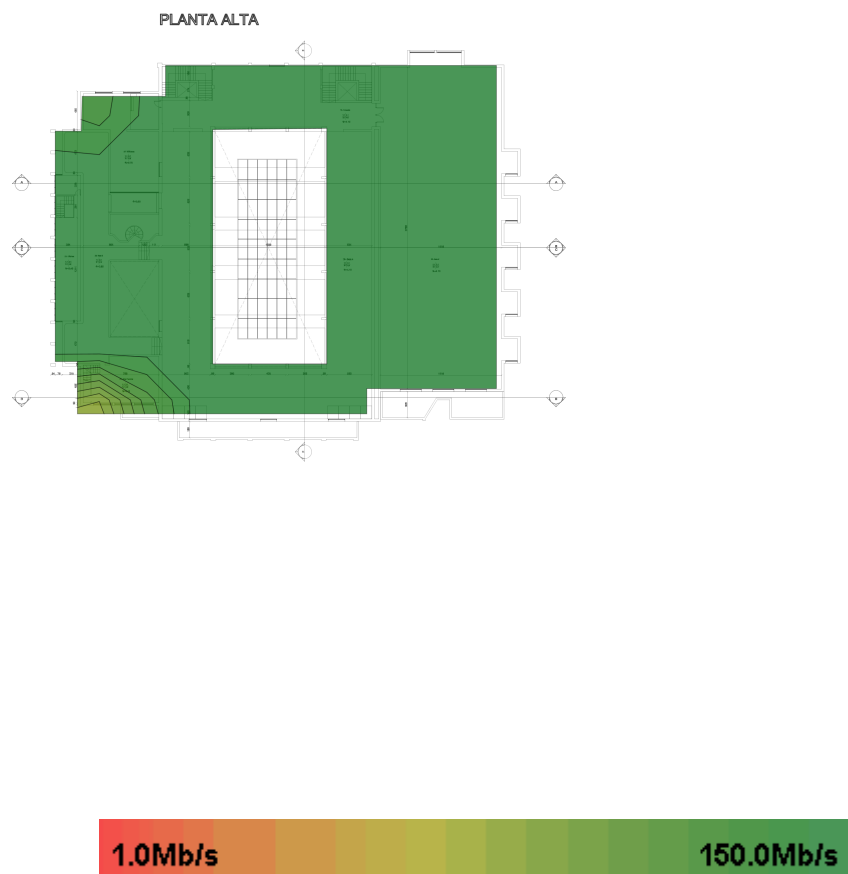
Number of Access Points indicates the number of access points audible at each location.



Wi-Fi Network Report

Data Rate for Planta_alta on 2.4 GHz band

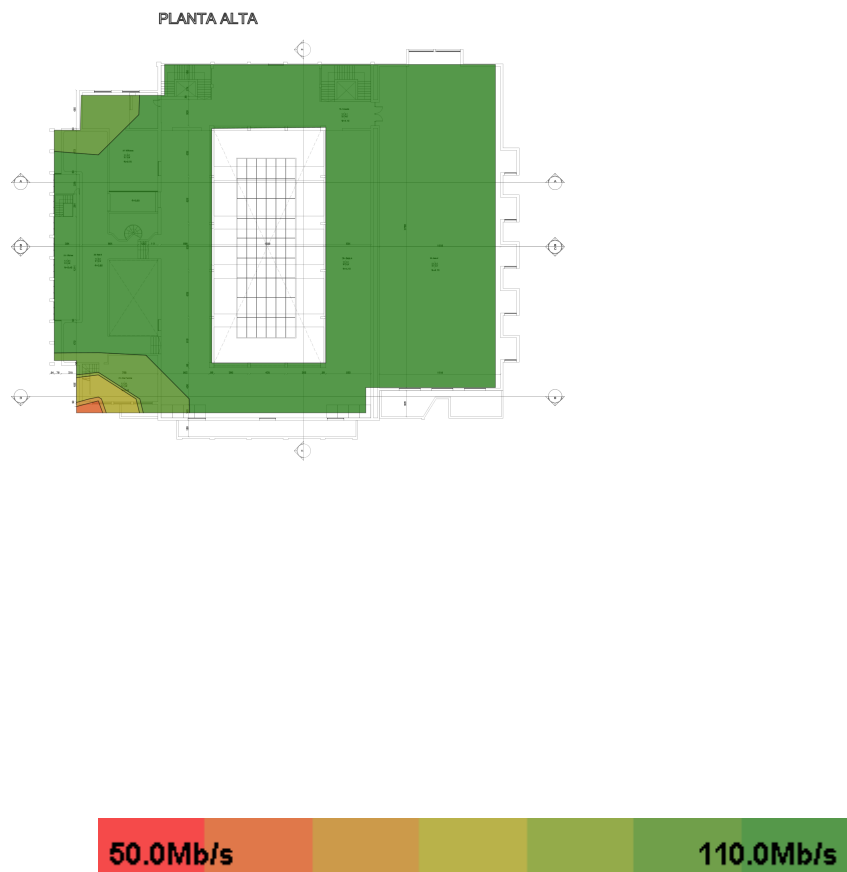
Data Rate is the highest possible speed (measured in megabits per second) at which the wireless devices will be transmitting data. Typically the true data throughput is about half of the data rate or less.



Wi-Fi Network Report

Throughput for Planta_alta on 2.4 GHz band

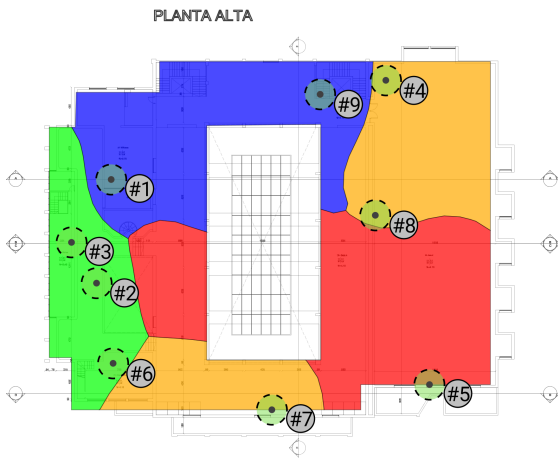
Displays the measured throughput. If no measured throughput is available, then the estimated maximum throughput is shown instead.



Wi-Fi Network Report

Associated Access Point for Planta_alta on 2.4 GHz band

Displays the access point the client device is associated with.
The image shows Predicted Association - Signal Strength



AP #	Access Point			
1	Cisco AP1602I (1)			
	 802.11n	11	60 mW	Cisco AP1602I 2.4GHz
	802.11n	136	0 mW	Cisco AP1602I 5GHz

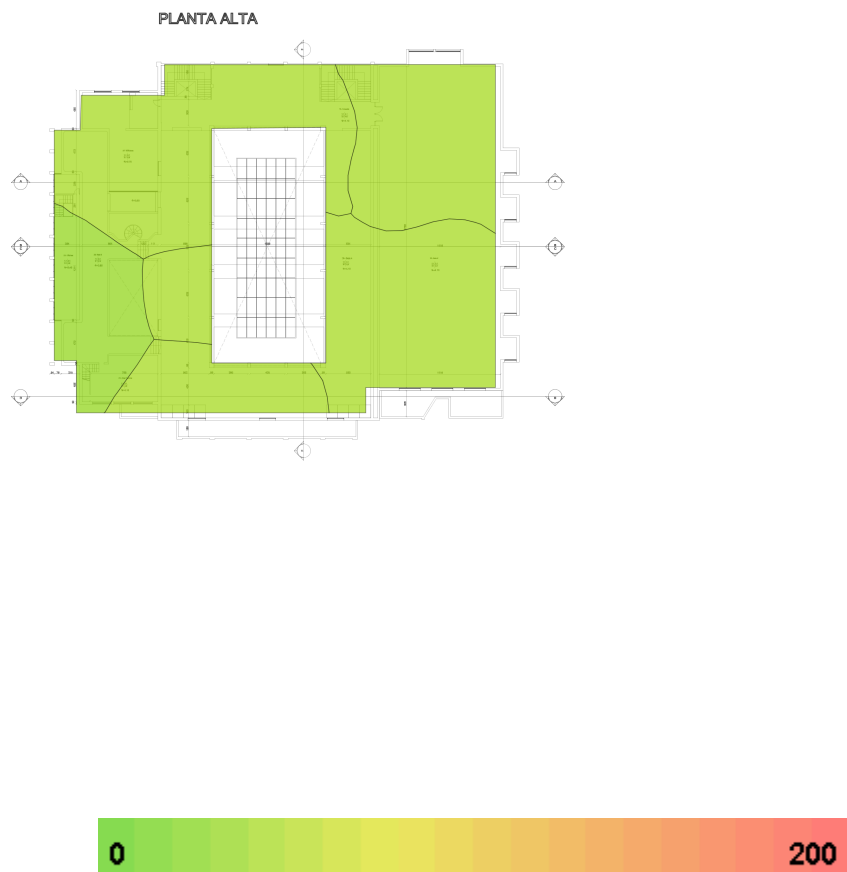
Wi-Fi Network Report

2	Cisco AP1602I (11)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	153	0 mW	Cisco AP1602I 5GHz
3	Cisco AP1602I (2)			
	 802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	44	0 mW	Cisco AP1602I 5GHz
4	Cisco AP1602I (3)			
	 802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	48	0 mW	Cisco AP1602I 5GHz
5	Cisco AP1602I (4)			
	 802.11n	6	60 mW	Cisco AP1602I 2.4GHz
	802.11n	120	0 mW	Cisco AP1602I 5GHz
6	Cisco AP1602I (5)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	112	0 mW	Cisco AP1602I 5GHz
7	Cisco AP1602I (6)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	157	0 mW	Cisco AP1602I 5GHz
8	Cisco AP1602I (7)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	100	0 mW	Cisco AP1602I 5GHz
9	Cisco AP1602I (8)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	128	0 mW	Cisco AP1602I 5GHz

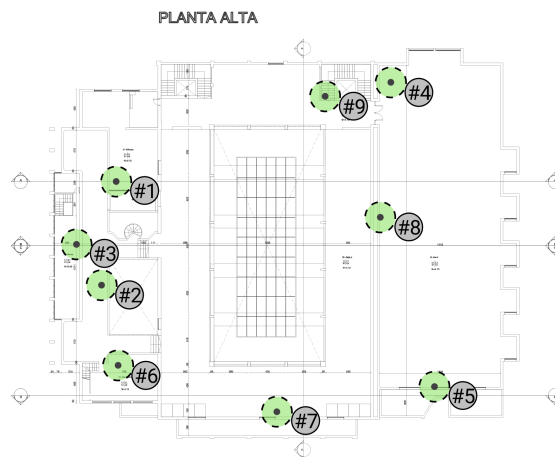
Wi-Fi Network Report

Capacity: Clients per AP for Planta_alta on 2.4 GHz band

Shows how the Wi-Fi clients configured in your Capacity Requirement are distributed between the access points. The image shows Requested Associations



Access Points on Planta_alta



My Access Points on Planta_alta

Simulated Access Points on Planta_alta

AP #	Access Point			
1	Cisco AP1602I (1)			
	802.11n	11	60 mW	Cisco AP1602I 2.4GHz
	802.11n	136	0 mW	Cisco AP1602I 5GHz
2	Cisco AP1602I (11)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	153	0 mW	Cisco AP1602I 5GHz
3	Cisco AP1602I (2)			
	802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	44	0 mW	Cisco AP1602I 5GHz
4	Cisco AP1602I (3)			
	802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	48	0 mW	Cisco AP1602I 5GHz
5	Cisco AP1602I (4)			
	802.11n	6	60 mW	Cisco AP1602I 2.4GHz
	802.11n	120	0 mW	Cisco AP1602I 5GHz
6	Cisco AP1602I (5)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	112	0 mW	Cisco AP1602I 5GHz
7	Cisco AP1602I (6)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	157	0 mW	Cisco AP1602I 5GHz
8	Cisco AP1602I (7)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	100	0 mW	Cisco AP1602I 5GHz
9	Cisco AP1602I (8)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	128	0 mW	Cisco AP1602I 5GHz

Measured Access Points on Planta_alta

None.

Other Access Points on Planta_alta

Simulated Access Points on Planta_alta

None.

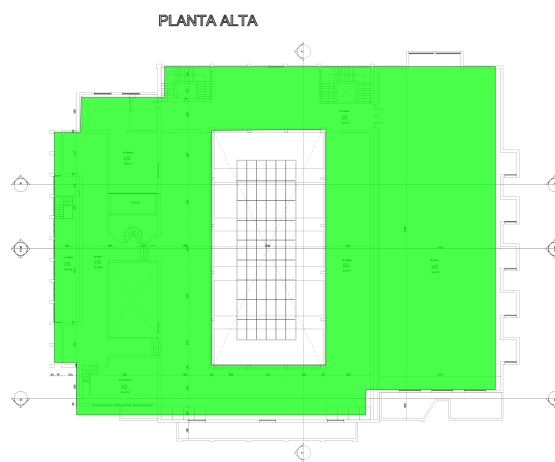
Measured Access Points on Planta_alta

None.

Wi-Fi Network Report

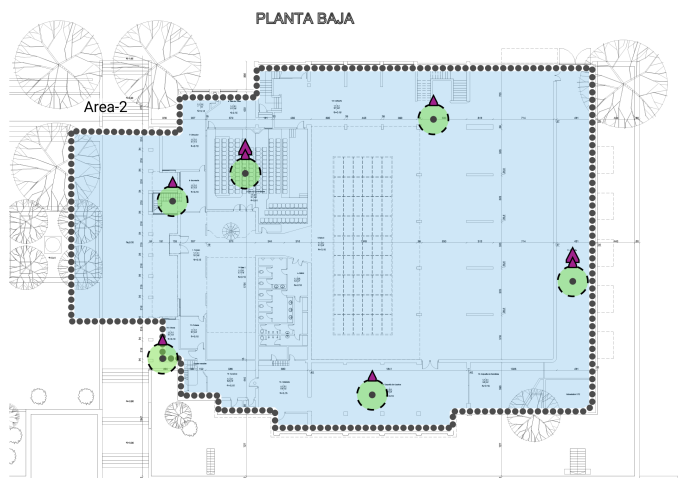
Channel Bandwidth for Planta_alta on 2.4 GHz band

Shows the maximum channel bandwidth available in each area.



Planta_baja

Survey routes and Access Points for Planta_baja



Area-2 (146.7 m²)

Coverage Requirement: Posifi	Signal Strength Min	-80.0 dBm
	Signal-to-noise Ratio Min	8.0 dB
	Data rate Min	2 Mbps
	Number of Access Points Min	5 at min. -80.0 dBm
	Channel Overlap Max	3 at min. -80.0 dBm
	Round Trip Time (RTT) Max	500ms
	Packet Loss Max	10.0 %

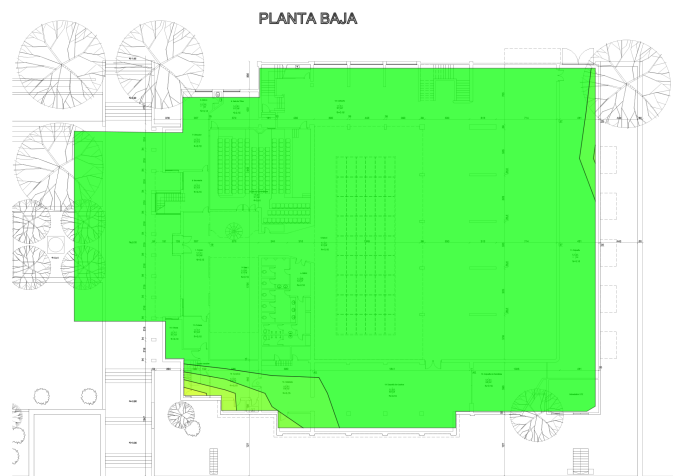
Wi-Fi Network Report

Capacity Requirement	100Generic Smartphone [Background Sync] Total: 100 (50 Mbits/s)
Notes	

Wi-Fi Network Report

Signal Strength for Planta_baja on 2.4 GHz band

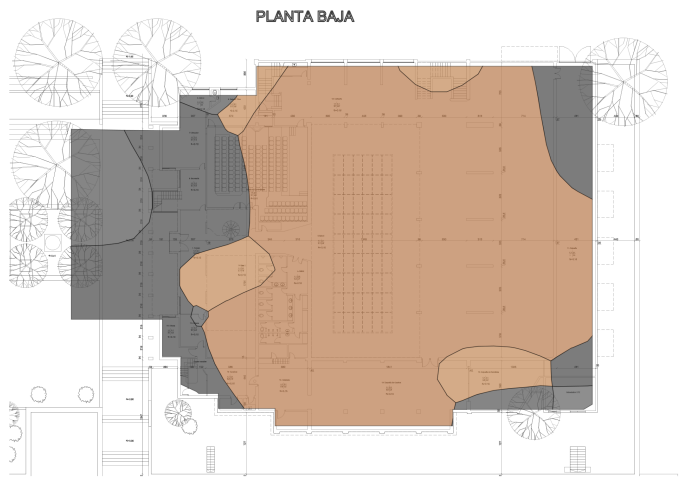
Signal Strength - sometimes called coverage - is the most basic requirement for a wireless network. As a general guideline, low signal strength means unreliable connections, and low data throughput.



Wi-Fi Network Report

Number of APs for Planta_baja on 2.4 GHz band

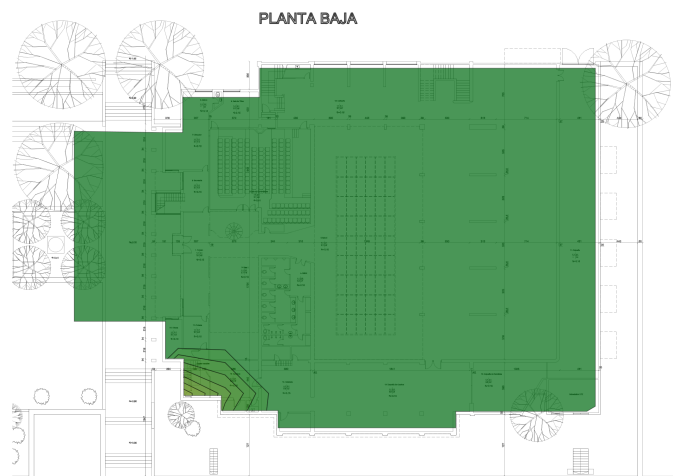
Number of Access Points indicates the number of access points audible at each location.



Wi-Fi Network Report

Data Rate for Planta_baja on 2.4 GHz band

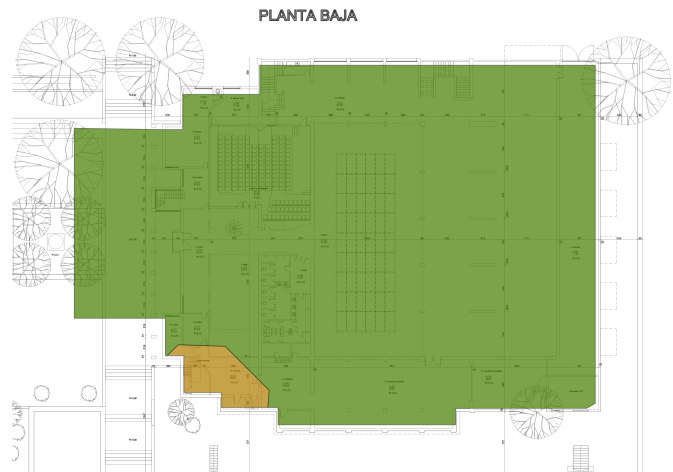
Data Rate is the highest possible speed (measured in megabits per second) at which the wireless devices will be transmitting data. Typically the true data throughput is about half of the data rate or less.



Wi-Fi Network Report

Throughput for Planta_baja on 2.4 GHz band

Displays the measured throughput. If no measured throughput is available, then the estimated maximum throughput is shown instead.



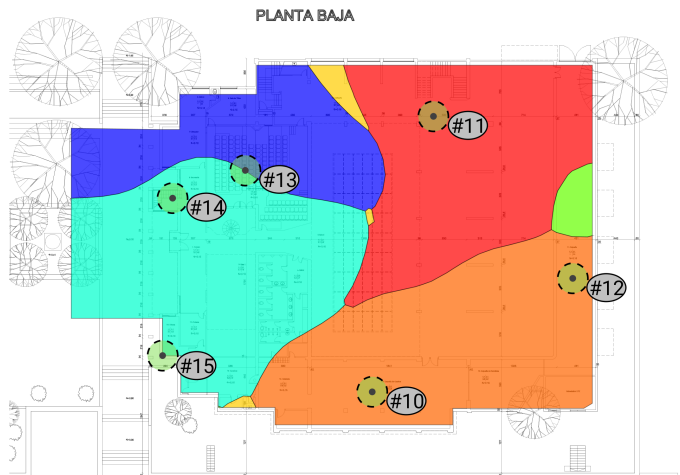
38.0Mb/s

110.0Mb/s

Wi-Fi Network Report

Associated Access Point for Planta_baja on 2.4 GHz band

Displays the access point the client device is associated with.
The image shows Predicted Association - Signal Strength



AP #	Access Point			
10	Cisco AP1602I (10)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	132	0 mW	Cisco AP1602I 5GHz

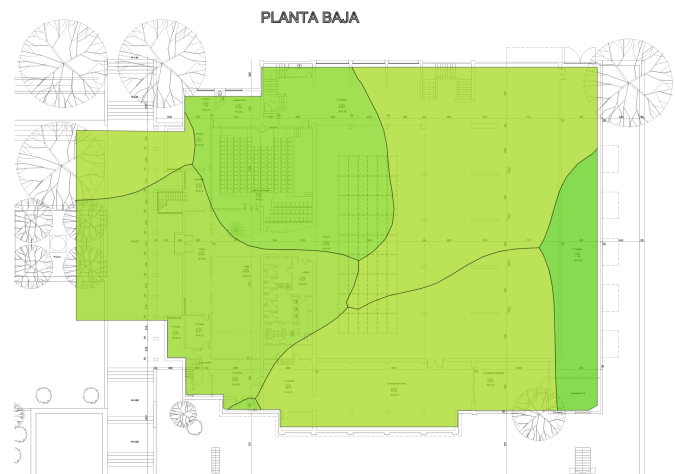
Wi-Fi Network Report

11	Cisco AP1602I (12)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	104	0 mW	Cisco AP1602I 5GHz
12	Cisco AP1602I (15)			
	 802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	56	0 mW	Cisco AP1602I 5GHz
13	Cisco AP1602I (16)			
	 802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	40	0 mW	Cisco AP1602I 5GHz
14	Cisco AP1602I (18)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	108	0 mW	Cisco AP1602I 5GHz
15	Cisco AP1602I (9)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	64	0 mW	Cisco AP1602I 5GHz

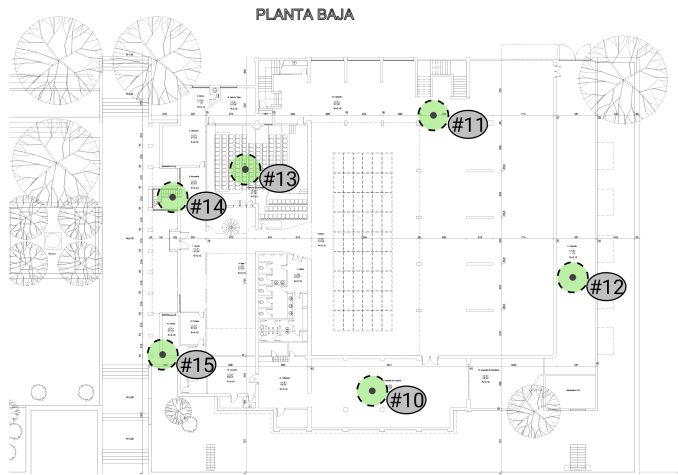
Wi-Fi Network Report

Capacity: Clients per AP for Planta_baja on 2.4 GHz band

Shows how the Wi-Fi clients configured in your Capacity Requirement are distributed between the access points. The image shows Requested Associations



Access Points on Planta_baja



Wi-Fi Network Report

My Access Points on Planta_baja

Simulated Access Points on Planta_baja

AP #	Access Point			
10	Cisco AP1602I (10)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	132	0 mW	Cisco AP1602I 5GHz
11	Cisco AP1602I (12)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	104	0 mW	Cisco AP1602I 5GHz
12	Cisco AP1602I (15)			
	802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	56	0 mW	Cisco AP1602I 5GHz
13	Cisco AP1602I (16)			
	802.11n	1	60 mW	Cisco AP1602I 2.4GHz
	802.11n	40	0 mW	Cisco AP1602I 5GHz
14	Cisco AP1602I (18)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	108	0 mW	Cisco AP1602I 5GHz
15	Cisco AP1602I (9)			
	Off	-	-	Cisco AP1602I 2.4GHz
	802.11n	64	0 mW	Cisco AP1602I 5GHz

Measured Access Points on Planta_baja

None.

Other Access Points on Planta_baja

Simulated Access Points on Planta_baja

None.

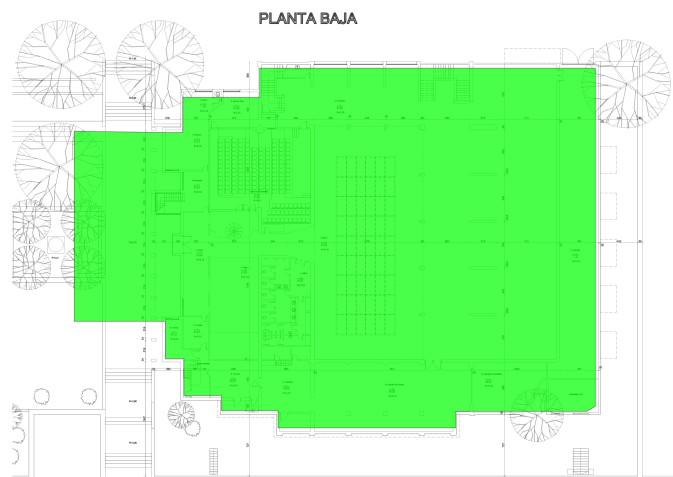
Measured Access Points on Planta_baja

None.

Wi-Fi Network Report

Channel Bandwidth for Planta_baja on 2.4 GHz band

Shows the maximum channel bandwidth available in each area.



Measured Access Points not placed on any map

My Access Points not placed on any map

None.

Other Access Points not placed on any map

None.

Referencias

- [1] Alisneaky. Support vector machines. <https://commons.wikimedia.org/w/index.php?curid=14941564>.
- [2] Alex E. Ball. Craft and folk art museum. self-guided tour mobile app developed for the craft and folk art museum. https://github.com/AlexEBall/Craft_and_Folk_Art_Museum_Mobile_App, 2018.
- [3] Cortes and Vapnik. Fingerprint kalman filter in indoor positioning applications. *Support-Vector Networks, Machine Learning, AT&T Bell Labs., Hohn-del, NJ 07733, USA.*, 20:273–297, 1995.
- [4] Davidson and Piché. A survey of selected indoor positioning methods for smartphones. *IEEE Communications surveys & tutorials*, 19(2), 2017.
- [5] Ministerio de diseño. La innovación en experiencias museísticas de la mano de ingeniería, tecnología y arte. <http://www.xn-ministeriodediseo-uxb.com/actualidad/la-innovacion-en-experiencias-museisticas-de-la-mano-de-ingenieria-tecnologia-y-arte/>, 2018.
- [6] The Warhol: Out Loud App development team. The warhol: Out loud app. <https://github.com/CMP-Studio/TheWarholOutLoud>.
- [7] Niklas Donges. The random forest algorithm. *machinelearning-blog.com*, 2018.
- [8] Ali-Löytty et al. Fingerprint kalman filter in indoor positioning applications. *3rd IEEE Multi-conference on Systems and Control, July 8-10, Saint Petersburg, RUSSIA (c)2009 IEEE*, pages 1678–1683, 2009.
- [9] Honkavirta et al. A comparative survey of wlan location fingerprinting methods. *6th Workshop on Positioning, Navigation and Communication (WPNC'09)*, page 243–251, 2009.
- [10] Devika Girish. Wi-fi vs bluetooth low energy (ibeacon) technology. *blog.beaconstac.com Software Engineer Blog*, 2014.
- [11] IEEE WG802.11 Wireless LAN Working Group. Telecommunications and information exchange between systems local and metropolitan area networks—specific requirements - part 11: Wireless lan medium access control (mac) and

Referencias

- physical layer (phy) specifications. *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pages 1–3534, 2016.
- [12] Ekahau Inc. ekahau site survey overview. <https://www.ekahau.com/products/ekahau-site-survey/overview/>, 2019.
 - [13] Facebook Inc. Learn the basics. <https://facebook.github.io/react-native/docs/tutorial>, 2019.
 - [14] Antonino Ingargiola and contributors. What is the jupyter notebook? https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html, 2015.
 - [15] Lui, Dempster, Gallagher, Li, and Rizos. Differences in rssi readings made by different wi-fi chipsets: A limitation of wlan localization. *International Conference on Localization and GNSS (ICL-GNSS)*, pages 53–57, 2011.
 - [16] Nils J. Nilsson. Introduction to machine learning an early draft of a proposed textbook. *Robotics Laboratory, Department of Computer Science, Stanford University, Stanford, CA 94305*, 1998.
 - [17] nobbot. ¿has probado wifi analyzer para optimizar la calidad de tu wifi? <https://www.nobbot.com/tecnologia/aplicaciones-moviles-tecnologia/wifi-analyzer-ventajas/>, 2018.
 - [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.
 - [19] Sebastian Raschka. *Python Machine Learning*. Packt Publishing Ltd., 2015.
 - [20] Amazon Web Services Sajee Mathew, Principal Solutions Architect. Overview of amazon web services. <https://d1.awsstatic.com/whitepapers/aws-overview.pdf>, 2018.
 - [21] Scikit-learn. Machine learning in python. <https://scikit-learn.org/stable/>, 2019.
 - [22] Bluetooth SIG. Bluetooth low energy, 2019.
 - [23] Cisco Systems. Cisco aironet 1600 series access point data sheet. <https://www.cisco.com/c/en/us/products/wireless/aironet-1600-series/index.html>.
 - [24] AnyPlace Development Team. A free and open indoor navigation service with superb accuracy! <https://anyplace.cs.ucy.ac.cy/>, 2018.
 - [25] FIND3 Development Team. Trilateration picture.

- [26] R. Schapire Y. Freund. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55:119–139, 1997.
- [27] W.J. Youden. Index for rating diagnostic tests. *Cancer*, 3:32–35, 1950.

Esta página ha sido intencionalmente dejada en blanco.

Índice de tablas

Esta página ha sido intencionalmente dejada en blanco.

Índice de figuras

2.1. Mientras el posicionamiento con Wi-Fi suele buscar obtener la posición exacta, en BLE se busca obtener la proximidad a cada beacon. Figura obtenida de [10].	6
2.2. Trilateración geométrica.	10
2.3. Trilateración a partir de medidas de <i>RSS</i>	10
2.4. Flujo básico de Sistema utilizando huellas.	11
2.5. Mapeo del sitio con valores de RSS representados en dBm [15]. . . .	12
2.6. Las máquinas de vectores de soporte mapean los vectores no linealmente separables en un espacio donde esto si sea posible [1].	14
2.7. Árbol de decisión. En cada uno de los nodos se realiza alguna prueba y los resultados de las mismas conducen a los nodos hoja que corresponden a la clasificación final.	14
2.8. Perceptrón [19].A partir de la diferencia entre la salida esperada y la salida del algoritmo se realiza una actualización de los pesos para la combinación lineal.	15
2.9. Perceptrón Multi-capa [19].La añadición de la capa intermedia le da una mayor capacidad y potencial para desarrollar problemas de mayor complejidad.	16
2.10. LDA y QDA utilizados en un mismo set de datos [18]. El levantar la suposición de tener las mismas matrices de covarianza genera que las fronteras de decisión sean cuadráticos y por tanto más flexibles.	18
2.11. Interfaz de aplicación de recolección de huellas de FIND3. Imagen de [25].	25
2.12. Tablero web de FIND3 [25].	25
3.1. Disposición de los APs en la planta baja del museo simulada en Ekahau.	29
3.2. Disposición de los APs en la planta alta del museo simulada en Ekahau.	30
3.3. Cantidad de APs visibles en cada punto de planta baja en la banda de 2.4GHz. Simulación a través de Ekahau.	31
3.4. Cantidad de APs visibles en cada punto de planta alta en la banda de 2.4GHz. Simulación a través de Ekahau.	32
3.5. Modelo de AP utilizado en la instalación del MNAV. Imagen de [23]	32
3.6. Distribución final de APs luego de la instalación.	34

Índice de figuras

3.7. Arquitectura del sistema implementado a partir de FIND3	35
3.8. Interfaz de aplicación de recolección de huellas nuevo.	38
3.9. Distribución de zonas en planta baja del museo. La zona 16 incluye el exterior del mismo. En rojo se muestra la distribución de las obras.	41
3.10. Distribución de zonas en planta alta del museo. En rojo se muestra la distribución de las obras.	42
3.11. Pantallas iniciales y menús principales de la aplicación MNAVegante.	43
3.12. Modo exhibición. La captura de pantalla despliega distintas imágenes con información correspondiente a las obras cercanas. Entrando mediante el botón inferior, se puede obtener información específica de la obra.	45
3.13. Modo exhibición. Información desplegada de la obra. Incluye texto descriptivo, la imagen y un audio reproducible donde se describe la obra.	45
3.14. Modo Recorrido a ciegas. Se despliega una imagen de la obra detectada, en este caso es una imagen genérica para marcar el inicio del recorrido. Se reproduce un audio inicial instructivo de cómo utilizar la aplicación.	48
3.15. Menú principal de aplicación web para información de la muestra. Aquí se carga la información de todas las obras a ser cargadas luego por la aplicación MNAVegante.	49
3.16. Campos para edición de información sobre cada obra en la aplicación web.	50
4.1. Matriz de confusión para el total de los datos	56
4.2. Tabla de exhaustividad y precisión para todos los datos. Entre paréntesis se anota la cantidad de muestras para cada zona.	57
4.3. Matriz de confusión para las MACs conocidas	59
4.4. Tabla de cobertura y precisión de las zonas para MACs conocidas .	60
4.5. Matriz de confusión de entrenamiento en doble banda y clasificación en banda de 2.4GHz	62
4.6. Matriz de confusión de entrenamiento y clasificación en banda de 2.4GHz	63
4.7. Tabla de cobertura y precisión de entrenamiento y clasificación en banda de 2.4GHz	64
4.8. Precisión en función de cantidad de APs utilizados en la instalación. Se muestra en cada punto el boxplot con las diez simulaciones por conjunto de APs sacados.	65
4.9. Precisión en función de cantidad de muestras. Se utiliza el boxplot para graficar las 10 simulaciones.	67
4.10. Distribución de zonas mínimas para estudio de resolución espacial.	69
4.11. Precisión general obtenida según la cantidad de zonas generadas en la distribución del espacio de planta baja.	70
5.1. Clasificaciones realizadas por día de la semana	74

5.2. Clasificaciones realizadas por día de la semana quitando los datos obtenidos el día de la apertura.	75
5.3. Clasificaciones realizadas por hora del día	76
5.4. Mapa con conteo de personas por zona.	77
5.5. Cantidad de descargas por usuarios únicos en el tiempo.	77
5.6. Bloqueos en el tiempo a partir de estadísticas de Google Play. . .	78

Esta es la última página.
Compilado el miércoles 1 mayo, 2019.
<http://iie.fing.edu.uy/>