



**UNIVERSIDAD DE CASTILLA-LA MANCHA  
ESCUELA SUPERIOR DE INFORMÁTICA**

**MÁSTER EN INGENIERÍA INFORMÁTICA**

**TRABAJO FIN DE MÁSTER**

Detección de comportamientos anómalos a través de dispositivos conectados en el hogar y su relación con las actividades de la vida diaria como medio para la mejora de la calidad de vida de personas mayores que viven solas

Adrián Sánchez-Miguel Ortega

Julio, 2024





**UNIVERSIDAD DE CASTILLA-LA MANCHA  
ESCUELA SUPERIOR DE INFORMÁTICA**

**Departamento de Tecnologías y  
Sistemas de Información**

**TRABAJO FIN DE MÁSTER**

**Detección de comportamientos anómalos a través de dispositivos conectados en el hogar y su relación con las actividades de la vida diaria como medio para la mejora de la calidad de vida de personas mayores que viven solas**

Autor: Adrián Sánchez-Miguel Ortega

Tutor(a): Jesús Fontecha Diezma

Co-tutor(a): Iván González Díaz

Julio, 2024



Detección de comportamientos anómalos y su relación con las actividades de la vida diaria  
© Adrián Sánchez-Miguel Ortega, 2024

Este documento se distribuye con licencia CC BY-NC-SA 4.0. El texto completo de la licencia puede obtenerse en <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

La copia y distribución de esta obra está permitida en todo el mundo, sin regalías y por cualquier medio, siempre que esta nota sea preservada. Se concede permiso para copiar y distribuir traducciones de este libro desde el español original a otro idioma, siempre que la traducción sea aprobada por el autor del libro y tanto el aviso de copyright como esta nota de permiso, sean preservados en todas las copias.

Este texto ha sido preparado con la plantilla  $\LaTeX$  para Trabajo Fin de Estudios en Ingeniería Informática para la UCLM publicada por [Jesús Salido](#) en el repositorio público Zenodo, DOI: [10.5281/zenodo.4561708](https://doi.org/10.5281/zenodo.4561708), como parte del curso « *$\LaTeX$  esencial para preparación de TFG, Tesis y otros documentos académicos*» impartido en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha. Si te ha resultado de utilidad te agradeceré que la cites [1] e incluyas en tus referencias como se indica en Zenodo con el DOI suministrado para todas las versiones.





TRIBUNAL:

Presidente: \_\_\_\_\_

Vocal: \_\_\_\_\_

Secretario(a): \_\_\_\_\_

FECHA DE DEFENSA: \_\_\_\_\_

CALIFICACIÓN: \_\_\_\_\_

PRESIDENTE

VOCAL

SECRETARIO(A)

Fdo.:

Fdo.:

Fdo.:



*A mis abuelos  
Por ser el mejor ejemplo a seguir*



## **Detección de comportamientos anómalos y su relación con las actividades de la vida diaria**

Adrián Sánchez-Miguel Ortega

Ciudad Real, Julio 2024

### **Resumen**

El presente trabajo se enmarca dentro del proyecto en curso financiado por el Ministerio de Ciencia e Innovación, en su convocatoria *Transición Ecológica y Digital 2021 (Ref. TED2021-130296A-I00)*, que tiene como principal objetivo mejorar la calidad de vida de las personas mayores que viven solas en sus hogares mediante el uso de nuevas tecnologías y procesos de intervención transparente al usuario.

Con ello, este Trabajo Fin de Máster pretende dar solución a la detección de comportamientos anómalos que tienen lugar en las rutinas diarias de un adulto mayor en el hogar, a través de la monitorización y análisis de los datos procedentes de dispositivos conectados distribuidos en el entorno, como elemento facilitador para cuidadores y profesionales, con el fin de mejorar la calidad de vida de la persona y conservar su autonomía el mayor tiempo posible.

Se desarrollará un sistema que, a través de la monitorización continuada de dispositivos conectados a electrodomésticos y otros aparatos electrónicos del hogar, permitirá tras preprocesamiento el análisis de los datos, teniendo en cuenta su correlación dentro del marco de las Actividades de la Vida Diaria. Se pretende encontrar patrones derivados de rutinas de actividades para, en un proceso posterior, y a través del empleo de técnicas de inteligencia artificial, detectar cambios anómalos que pudieran ser indicativos de posibles deterioros cognitivos u otros problemas relacionados, que afecten negativamente a la calidad de vida de las personas mayores en el hogar, o a su propia autonomía.



## **Detection of anomalous behavior and their relationship with Activities of Daily Living**

Adrián Sánchez-Miguel Ortega  
Ciudad Real, July 2024

### **Abstract**

This work is part of the ongoing project funded by the Ministry of Science and Innovation, in its call *Ecological and Digital Transition 2021 (Ref. TED2021-130296A-I00)*, whose main objective is to improve the quality of life of older people living alone at home through the use of new technologies and transparent intervention processes to the user.

With this, this Master Thesis aims to provide a solution to the detection of anomalous behaviors that take place in the daily routines of an elderly person at home, through the monitoring and analysis of data from connected devices distributed in the environment, as a facilitating element for caregivers and professionals, in order to improve the quality of life of the person and preserve their autonomy as long as possible. A system will be developed that will collect data and perform continuous monitoring from devices connected to different appliances distributed in the home.

A system will be developed which, through the continuous monitoring of devices connected to household appliances and other electronic devices in the home, will allow the analysis of the data after pre-processing, taking into account their correlation within the framework of Activities of Daily Living. The aim is to find patterns derived from activity routines in order to, in a subsequent process, and through the use of artificial intelligence techniques, detect anomalous changes that could be indicative of possible cognitive deterioration or other related problems, that can negatively affect the quality of life of elderly people at home, or their own autonomy.



# Agradecimientos

---

Al igual que en los agradecimientos del TFG, debo empezar de nuevo mencionando a mis padres, Ángel y Alicia, los cuales no solo me dieron la vida, sino que ahora que vivo lejos me dan media hora de su tiempo cuando los llamo para preguntarles qué tal les ha ido el día todos los días de la semana.

Mencionar también a mis hermanos Ángel y Alicia, que cada que nos reencontramos en casa y nos echamos una Wii, volvemos a sentir que tenemos 10 años, y a mi hermano Álvaro, que está viviendo su mejor vida, lo cual me hace muy feliz.

A mi querida Elena, por soportar que me quede hasta las tantas con el TFM en vez de ver cualquier serie con ella, y darle las gracias por ser el pilar fundamental en el cual seguir apoyándome en mi día a día.

A mis abuelos, que seguro estarán orgullosos de mí y de mis hermanos, allá donde estén, y al resto de mis tíos, tías y familia por animarme desde cualquier punto de la geografía española a continuar haciendo aquello que más me gusta.

A mis amigos del pueblo, que los quiero mucho más de lo que los veo últimamente, y que me recuerdan de donde vengo y siempre me dan momentos increíbles.

A la gente del Computeam, por hacerme desconectar todos los miércoles a las 20 horas de la tarde a base de pelotazos a la redonda y muchas risas.

Por supuesto, también agradecer a mis tutores Jesús e Iván, por su paciencia y amabilidad a la hora de realizar el proyecto TAICare, y al resto del grupo MAmI con el que he trabajado estos meses: David, Laura, Cosmin (y Gabi), Ramón, Luis, Tania, Esperanza, Alfonso, Bea, Ale, Celia, Yulith, Guille, Marcos, y al director del grupo de Investigación, José Bravo, por invitarme al congreso UCAmI del año pasado.

*Adrián Sánchez-Miguel Ortega*  
Ciudad Real, 2024



# Acrónimos

---

## LISTA DE ACRÓNIMOS

Lista con los acrónimos empleados en el documento ordenados alfabéticamente.

AAL	: Ambient Assisted Living
API	: Application Programming Interface
ADL	: Activity of Daily Living
BADL	: Basic Activity of Daily Living
CPU	: Central Processing Unit
CSV	: Comma-Separated Values
CUSUM	: Cumulative Sum Control Chart
DBSCAN	: Density-Based Spatial Clustering
ESI	: Escuela Superior de Informática
EE.UU.	: Estados Unidos
GMM	: Gaussian Mixture Modelling
GPU	: Graphic Processing Unit
HMM	: Modelo oculto de Markov
HTTP	: Protocolo de transferencia de hipertexto
IA	: Inteligencia Artificial
IADL	: Instrumental Activity of Daily Living
ILM	: Intrusive Load Monitoring
IoT	: Internet de las Cosas
JSON	: JavaScript Object Notation
LSTM	: Long Short-Term Memory
MAMi	: Modelling Ambient Intelligence
MQTT	: Message Queuing Telemetry Transport
NILM	: Non-Intrusive Load Monitoring
NN	: Neural Network
PIR	: Sensores infrarrojos pasivos
PR	: Prototipado Rápido
RAD	: Rapid Application Development
REDD	: Reference Energy Disaggregation Data Set
RFID	: Etiquetas de identificación por radiofrecuencia
RU	: Reino Unido
SiSHoDit	: Simulator of Smart Houses Digital Twins
SQL	: Structured Query Language
SDLC	: Software Development Life Cycle
SVM	: Árbol de decisión
TFG	: Trabajo Fin de Grado
TFM	: Trabajo Fin de Máster
TPU	: Tensor Processing Unit
UCAmI	: Ubiquitous Computing and Ambient Intelligence

UCLM : Universidad de Castilla-La Mancha  
UK-DALE : UK Domestic Appliance-Level Electricity

# Índice general

---

<b>Resumen</b>	<b>V</b>
<b>Abstract</b>	<b>VII</b>
<b>Agradecimientos</b>	<b>IX</b>
<b>Notación y acrónimos</b>	<b>XI</b>
<b>Índice de figuras</b>	<b>XV</b>
<b>Índice de tablas</b>	<b>XVII</b>
<b>Índice de listados</b>	<b>XIX</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Justificación de la necesidad . . . . .	2
1.3. Descripción general . . . . .	2
1.4. Estructura del documento . . . . .	3
<b>2. Objetivo</b>	<b>5</b>
2.1. Objetivo general . . . . .	5
2.2. Objetivos específicos . . . . .	5
<b>3. Estado del arte</b>	<b>7</b>
3.1. Actividades de la Vida Diaria . . . . .	7
3.2. Monitorización Intrusiva vs. Monitorización No Intrusiva . . . . .	8
3.3. ILM y NILM: Enfoques Específicos en Monitorización de Cargas Eléctricas . . . . .	9
3.4. Detección de Actividades de la Vida Diaria . . . . .	10
3.5. Detección de Anomalías . . . . .	11
<b>4. Metodología</b>	<b>13</b>
4.1. Metodología de desarrollo software . . . . .	14
4.2. Primer Prototipo: Reporte técnico tras estudio para la Elección de Enchufes Inteligentes	15
4.3. Segundo Prototipo: Modelo para la extracción de Información de los Enchufes . . . . .	15
4.4. Tercer Prototipo: Modelo para la detección de IADLs a través de Machine Learning	15
4.5. Cuarto Prototipo: Componente software Rust para la Recogida Automática de Información . . . . .	16
4.6. Quinto Prototipo: Modelo para la detección de Anomalías en IADLs . . . . .	16
4.7. Recursos para el desarrollo . . . . .	17

---

<b>5. Resultados</b>	<b>21</b>
5.1. Planificación previa . . . . .	21
5.2. Iteración 1: Reporte técnico tras estudio para la elección de enchufes inteligentes . . . . .	21
5.3. Iteración 2: Modelo para la extracción de información de los enchufes . . . . .	23
5.4. Iteración 3: Modelo para la detección de IADLs . . . . .	25
5.5. Iteración 4: Componente software Rust para la recogida de información . . . . .	32
5.6. Iteración 5: Modelo para la detección de anomalías en la IADLs . . . . .	37
<b>6. Conclusiones</b>	<b>43</b>
6.1. Discusión . . . . .	43
6.2. Objetivos alcanzados . . . . .	44
6.3. Trabajo futuro . . . . .	45
6.4. Justificación de competencias adquiridas . . . . .	45
6.5. Agradecimientos . . . . .	45
<b>Bibliografía</b>	<b>47</b>
<b>A. Repositorios con el código del proyecto</b>	<b>55</b>

# Índice de figuras

---

1.1. Pirámides de población en España y proyección para 2052 . . . . .	1
1.2. Porcentaje de personas que viven solas en España según edad y sexo año 2020. . . . .	2
1.3. Resumen del trabajo por etapas . . . . .	3
3.1. Métodos de monitorización no intrusiva . . . . .	9
3.2. Métodos de monitorización intrusiva . . . . .	9
4.1. Fases del proyecto . . . . .	13
4.2. Diagrama de Gantt del proyecto . . . . .	14
4.3. Metodología PR aplicada al proyecto . . . . .	14
4.4. Resumen de los prototipos y sus resultados . . . . .	17
4.5. Enchufe Tapo P110 . . . . .	19
4.6. Raspberry Pi 5 en su carcasa . . . . .	20
5.1. Simulación de HS110 . . . . .	22
5.2. Información en Node-RED . . . . .	23
5.3. Diagrama de secuencia del trabajo de investigación . . . . .	24
5.4. Marco de adquisición ( <i>framework</i> ) del proyecto . . . . .	24
5.5. Información en Grafana . . . . .	25
5.6. Serie temporal con el gasto de varios electrodomésticos a lo largo del día . . . . .	26
5.7. Aplicación y correlación del algoritmo K-Means para con los datos de electrodomésticos . . . . .	27
5.8. Estructura de una célula en una red LSTM . . . . .	28
5.9. Capas de la red LSTM del proyecto . . . . .	29
5.10. Precisión de la red neuronal LSTM . . . . .	30
5.11. Comparación los algoritmos de K-Means, GMM y clustering jerárquico . . . . .	31
5.12. Matriz de correlación de los algoritmos . . . . .	32
5.13. Diagrama de flujo de la API . . . . .	33
5.14. Esquema de la base de datos . . . . .	35
5.15. Raspberry recogiendo datos de los enchufes inteligentes . . . . .	37
5.16. Herramienta SiSHoDit en funcionamiento . . . . .	38
5.17. Comparación matriz hogar sintético con hogar del REFIT . . . . .	39
5.18. Valores anómalos en la actividad a lo largo de un mes . . . . .	41



# Índice de tablas

---

3.1. BADLs de los distintos índices . . . . .	8
5.1. Comparación de <i>datasets</i> públicos . . . . .	26
5.2. Comparación de los tres algoritmos no supervisados . . . . .	31
6.1. Competiciones del Máster de Ingeniería Informática . . . . .	46



# Índice de listados

---

5.1. Hiperparametrización del algoritmo K-Means . . . . .	28
5.2. Red neuronal LSTM . . . . .	29
5.3. Código para la conexión con los enchufes . . . . .	33
5.4. Formato de logging de la aplicación SiSHoDit . . . . .	39

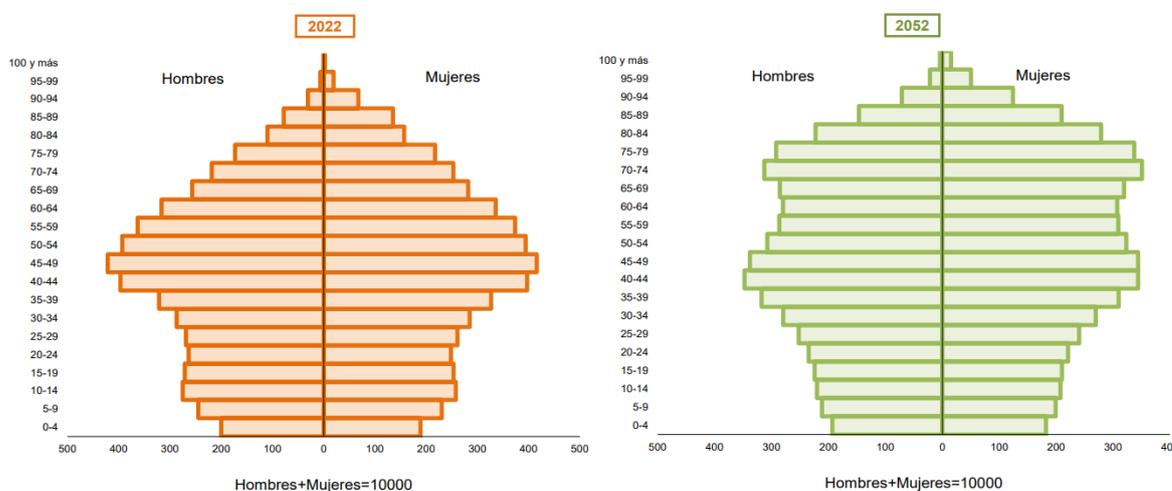


# Introducción

En este capítulo se aborda la motivación del trabajo. Se señala la necesidad que lo origina, la propuesta para aportar a ello, y la estructura del documento que aquí se presenta.

## 1.1. MOTIVACIÓN

El envejecimiento de la población se ha convertido en un fenómeno cada vez más evidente, y su impacto en los servicios sanitarios públicos puede llegar a generar importantes desafíos [2]. Según la Fig. 1.1, la población de edad avanzada está experimentando un crecimiento notable [3]. Este fenómeno plantea interrogantes sobre la capacidad de los servicios de salud para satisfacer las crecientes demandas de una población que envejece.



**Figura 1.1:** Pirámides de población en España y proyección para 2052

Un aspecto destacado, como se refleja en la Fig. 1.2, es la creciente tendencia de las personas mayores a vivir solas en sus hogares [4]. Este cambio en la dinámica familiar indica una preferencia por la independencia y la autonomía en la tercera edad. Cada vez más ancianos aspiran a prolongar su vida en un entorno hogareño sin depender en gran medida de cuidadores o asistencia externa, o al menos minimizando su necesidad.

Este deseo de autonomía plantea importantes desafíos para los responsables de la planificación de servicios de salud y sociales. Destaca la importancia de abordar esta creciente demanda de una manera que permita a los ancianos vivir de forma independiente y realizar acciones cotidianas de la vida diaria mientras se garantiza la disponibilidad de servicios adecuados. Por tanto, la sociedad se enfrenta al reto de adaptar y mejorar los recursos y programas existentes.

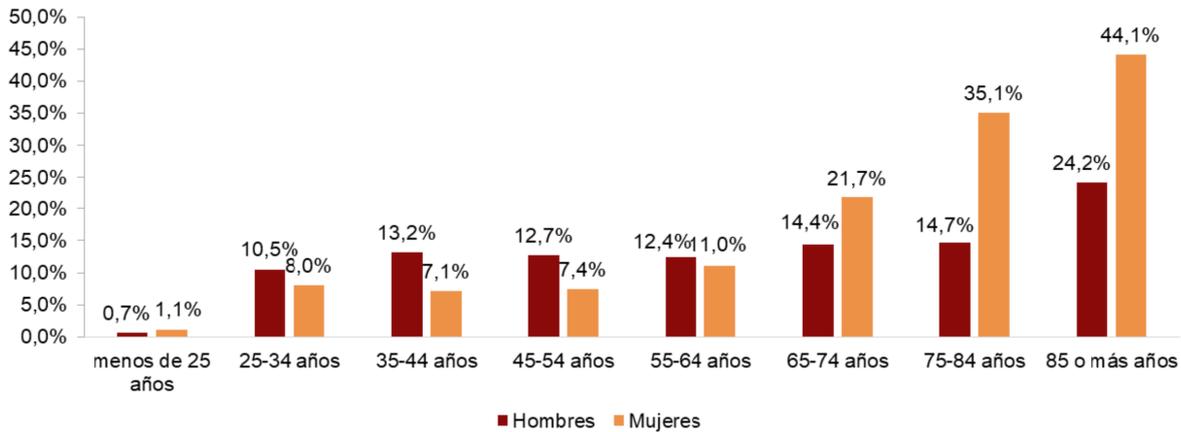


Figura 1.2: Porcentaje de personas que viven solas en España según edad y sexo año 2020.

## 1.2. JUSTIFICACIÓN DE LA NECESIDAD

En respuesta al desafío de mantener la independencia de las personas mayores, se están implementando innovaciones tecnológicas que buscan mejorar la calidad de vida de los ancianos. Una estrategia prometedora es el monitoreo de las **Actividades de la Vida Diaria** (*Activities of Daily Living*, ADLs). Las ADLs abarcan una gama de actividades básicas y necesarias para el funcionamiento independiente en la vida cotidiana [5]. Estas actividades se dividen comúnmente en dos categorías: actividades básicas de la vida diaria (BADLs) y actividades instrumentales de la vida diaria (IADLs). Mientras que las BADLs incluyen tareas fundamentales como comer, vestirse y asearse [6], las IADLs comprenden actividades más complejas relacionadas con la gestión del hogar y la relación con el entorno, normalmente. Entre estas últimas se encuentran la preparación de comidas, el transporte, las compras y la gestión de la medicación.

En este proyecto de investigación, mediante el uso de enchufes inteligentes, los cuales se conectarán a electrodomésticos que se puedan asociar a actividades instrumentales, se pretende poder monitorizar las acciones diarias llevadas a cabo. Al asociarse con electrodomésticos clave utilizados en estas, estos dispositivos recopilan datos sobre el uso a lo largo del tiempo. Esto permitirá detectar, aparte de las actividades ya mencionadas, posibles problemas o desviaciones.

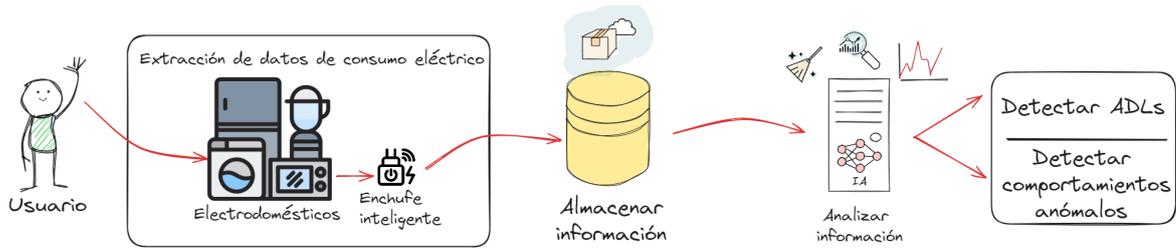
Este enfoque tecnológico no solo proporciona información sobre el bienestar y la autonomía de los ancianos, sino que también puede mejorar la eficiencia de los servicios de atención médica al permitir intervenciones preventivas. Sin embargo, es crucial abordar las preocupaciones éticas y de privacidad asociadas con la recopilación de datos personales, asegurando que estas tecnologías se utilicen de manera ética y segura.

Por tanto, la integración de enchufes inteligentes para monitorear las IADLs representa un avance en la búsqueda de soluciones tecnológicas para apoyar la independencia de los ancianos. Este enfoque no solo tiene el potencial de mejorar la calidad de vida de la población en envejecimiento, sino que también ofrece oportunidades para una atención más eficiente y preventiva.

## 1.3. DESCRIPCIÓN GENERAL

Uno de los pilares de este trabajo será el deducir las actividades cotidianas de las personas mayores que residen solas a través del análisis de datos sobre el consumo energético con el fin de determinar, en última instancia, comportamientos anormales que puedan derivar en problemas de salud, afectando a la calidad de vida de las personas.

Como se puede ver en la Fig. 1.3, la propuesta se divide en varias etapas, empezando por la recolección de esta información, para la cual se emplearán enchufes inteligentes, conectados a electrodomésticos relevantes. Esta información será almacenada eligiendo el tipo de base de datos



**Figura 1.3:** Resumen del trabajo por etapas

que mejor se adapte al trabajo, y el procesamiento de los datos se llevará a cabo mediante el uso de técnicas de aprendizaje automático. Se procede a explicar ahora con un poco más de detalle cada una de las etapas de este trabajo de investigación:

### 1.3.1. Extracción de datos de consumo eléctrico

El proceso de recolección de datos implica el uso de un dispositivo electrónico que haga de intermediario para recoger la información de consumo de energía de los enchufes inteligentes, requiriendo periódicamente estos datos a todos los enchufes conectados. Esta configuración permite una recolección de datos precisa y en tiempo real, proporcionando un conjunto de datos completo para el análisis.

### 1.3.2. Almacenamiento de Datos

Los datos recogidos se enviarán a una base de datos, donde se almacenarán en varias colecciones para su fácil recuperación y organización. El sistema de almacenamiento elegido asegurará la integridad y accesibilidad de los datos.

### 1.3.3. Análisis de Datos

Los datos almacenados, una vez filtrados, se recuperarán para crear conjuntos de datos que se analicen utilizando técnicas de aprendizaje automático. Estas técnicas podrán ser tanto de aprendizaje supervisado como no supervisado, midiendo su rendimiento con la información proporcionada, para identificar patrones y comportamientos anómalos en las actividades diarias de las personas mayores.

### 1.3.4. Resultados Esperados

A través de este trabajo, esperamos identificar y predecir con precisión las actividades diarias de las personas mayores que viven solas, al igual que las posibles desviaciones en los patrones resultantes. Esto podría llevar al desarrollo de sistemas inteligentes que pueden proporcionar asistencia oportuna y mejorar la calidad de vida de las personas mayores.

## 1.4. ESTRUCTURA DEL DOCUMENTO

El TFM constará de las siguientes secciones:

1. **Introducción.** Se responde a la pregunta sobre la motivación o pertinencia de este trabajo, se da una justificación del mismo y se establece la estructura del documento.
2. **Objetivos.** Se determina de modo claro el propósito general del trabajo descrito y se desglosa en objetivos parciales y requisitos que deberá cumplir el trabajo.
3. **Estado del arte.** En esta sección se abordarán conceptos fundamentales de las ADL e IADL. Aparte, se realizará una revisión exhaustiva de los estudios más destacados en el campo de la monitorización de estas actividades, nombrando las tecnologías utilizadas para ello, y de los comportamientos anómalos que pueden surgir en ellas.

4. **Metodología.** Se establece la metodología de trabajo y se determina la planificación inicial del trabajo.
5. **Resultados.** Se reportan los resultados obtenidos tras el análisis e implementación de cada una de las iteraciones que forman el desarrollo de este trabajo.
6. **Conclusiones.** Se extraen conclusiones tras la realización del trabajo, se justifican las competencias aprendidas y se aportan mejoras de futuro en la investigación de este trabajo.

---

## CAPÍTULO 2

# Objetivo

---

En este capítulo se exponen los objetivos y subobjetivos que este proyecto espera cumplir.

### 2.1. OBJETIVO GENERAL

El objetivo general de este trabajo es la detección de comportamientos anómalos en personas mayores que viven solas en el hogar utilizando dispositivos inteligentes conectados, a través de la identificación de las Actividades de la Vida Diaria que realizan en casa, facilitando así la identificación temprana de problemas que impacten negativamente en su autonomía y su calidad de vida.

La detección de comportamientos anómalos se abordará a través de la monitorización detallada de las interacciones de los usuarios con aparatos conectados a estos dispositivos. La recolección de datos se llevará a cabo de manera no intrusiva, utilizando tecnologías como enchufes inteligentes para capturar patrones de comportamiento en actividades cotidianas. La relación con las ADLs será esencial, ya que se busca entender cómo las rutinas diarias y las interacciones con el entorno pueden llegar a insinuar posibles deterioros cognitivos.

La integración de algoritmos de Machine Learning, tanto supervisados como no supervisados, permitirá analizar los datos recopilados e identificar patrones característicos de comportamiento a través de la detección de estas ADLs. Este enfoque permitirá no solo establecer rutinas, sino también crear un marco teórico que permita definir qué serán las desviaciones en los datos que podrían indicar situaciones anómalas.

### 2.2. OBJETIVOS ESPECÍFICOS

Para conseguir el objetivo general, se debe hacer una división de este en varios objetivos:

#### 2.2.1. Diseñar un marco de adquisición de datos para el establecimiento de patrones ad-hoc, relacionados con rutinas y comportamientos característicos

Este objetivo parcial se centra en desarrollar un marco para la adquisición de datos a través de enchufes inteligentes, con el fin de identificar comportamientos anómalos basándose en las rutinas diarias. La esencia del proyecto radica en capturar, procesar y analizar información de las actividades domésticas, utilizando dichos dispositivos como nodos de información. El diseño de este sistema debe ser lo suficientemente sencillo para que su instalación no represente una barrera para los usuarios.

#### 2.2.2. Monitorizar las interacciones del usuario con los dispositivos conectados en el entorno (monitorización de bajo nivel), incluyendo la adquisición, comunicación y preprocesamiento de los datos, y su relación con el marco propuesto

Este objetivo parcial se centra en la implementación de un sistema de monitorización que capte detalladamente las interacciones del usuario con dispositivos en un entorno conectado, abordando la adquisición, comunicación, preprocesamiento e integración de datos en el marco ya diseñado. La monitorización de bajo nivel incluirá mecanismos no intrusivos para capturar datos en tiempo real

y con precisión, cubriendo todos los dispositivos conectados para obtener una visión completa del entorno doméstico.

El sistema asegurará una comunicación eficiente entre dispositivos y el *gateway* o mediante protocolos que minimicen pérdidas de información y garanticen la integridad de los datos. El preprocesamiento será crucial para preparar la información para el posterior análisis, incluyendo el manejo de inconsistencias, seleccionando solo atributos relevantes para alimentar los futuros algoritmos de detección de anomalías. Estos pasos optimizarán el uso de recursos y mejorarán la precisión del análisis. La correcta integración de estos datos permitirá una comprensión completa de las actividades del usuario, facilitando el reconocimiento de actividades y comportamientos anómalos, y asegurando que la monitorización.

### **2.2.3. Extraer conocimiento de los datos adquiridos (de datos públicos relacionados, o de datos sintéticos), mediante el empleo de algoritmos *Machine Learning* de aprendizaje supervisado y no supervisado, tanto para el análisis de las rutinas comportamentales como para la detección de situaciones anómalas, y su relación con las Actividades de la Vida Diaria**

Este objetivo parcial se centra en la extracción de conocimiento a partir de los datos adquiridos, utilizando tanto algoritmos de *Machine Learning* supervisados como no supervisados para analizar rutinas de actividades y para detectar situaciones anómalas. Los algoritmos no supervisados permitirán identificar automáticamente grupos de datos similares que representen diversas actividades o patrones de comportamiento sin necesidad de etiquetas previas, lo cual es esencial para la flexibilidad del sistema, ya que este se utilizará en distintos hogares, y los comportamientos variarán significativamente. En contraste, los algoritmos supervisados se utilizarán para entrenar el sistema con los datos etiquetados, lo que facilitará el reconocimiento de patrones específicos relacionados con comportamientos habituales. La efectividad de estos procesos dependerá de la elección adecuada de características relevantes y la construcción de un conjunto de datos de entrenamiento representativo.

### **2.2.4. Realizar una evaluación empírica, mediante la aproximación del escenario a un entorno real**

Este objetivo parcial se enfoca en la evaluación empírica del sistema, buscando validar su eficacia y viabilidad práctica en un escenario que simula condiciones reales de vida cotidiana. Para ello, se implementará el sistema en un entorno que imita las diversas situaciones y actividades diarias de los usuarios, utilizando enchufes inteligentes y dispositivos conectados. Este objetivo observar y medir el desempeño del sistema en condiciones dinámicas, recopilando datos en tiempo real y evaluando su capacidad para reconocer actividades y adaptarse a cambios inesperados.

La elección estratégica de este entorno de prueba y el análisis continuo durante la fase de evaluación son cruciales para asegurar no solo la eficacia técnica del sistema, sino también su impacto en la vida diaria de los usuarios. La precisión de las predicciones y la adaptabilidad del sistema serán examinadas, para así identificar posibles mejoras y ajustes que optimicen la integración de la tecnología en el día a día de las personas a quienes está destinado. Esta fase es fundamental para garantizar que el proyecto sea práctico, útil y bien recibido.

## Estado del arte

---

Los entornos inteligentes, potenciados por el Internet de las cosas (IoT), han emergido como soluciones innovadoras destinadas a mejorar la calidad de vida y la seguridad de las personas mayores en sus propios hogares [7, 8]. La capacidad de recopilar datos de manera eficiente y aplicar análisis avanzados ha dado lugar a nuevas posibilidades para monitorizar la salud, el bienestar y las rutinas diarias de este segmento de la población.

En este contexto, el presente proyecto se sitúa en la intersección de estos enfoques, explorando la viabilidad y eficacia de la monitorización mediante enchufes inteligentes en hogares de personas mayores. Estos dispositivos, al ser elementos cotidianos y de fácil integración, proporcionan una solución práctica y no intrusiva para comprender las rutinas y comportamientos característicos de este grupo demográfico.

Este análisis del estado del arte no solo busca contextualizar el trabajo presente, sino también destacar las tendencias emergentes y las áreas más exploradas. La convergencia de la monitorización IoT, presenta una oportunidad única para avanzar hacia entornos inteligentes más adaptables y centrados en las necesidades de las personas mayores [9].

A medida que exploramos los estudios previos, es evidente que el camino hacia la creación de entornos inteligentes efectivos en hogares de personas mayores es dinámico y en constante evolución. Este proyecto contribuye al tejido de conocimiento existente, aportando nuevas perspectivas y soluciones que podrían dar forma a futuras investigaciones y, en última instancia, mejorar la calidad de vida de aquellos que residen en la comodidad de sus hogares.

### 3.1. ACTIVIDADES DE LA VIDA DIARIA

Para saber de manera concisa cuáles deben ser las Actividades de la Vida Diaria, Básicas o Instrumentales (BADL o IADL) que se pueden o deben monitorizar en entornos inteligentes, es necesario un estudio de las mismas, ya que se pueden encontrar distintos índices clasificatorios de estas actividades, aparte del original citado previamente.[5]. En el índice original *de Actividades de la Vida Diaria* de Katz, se creó una herramienta sistemática para medir 6 actividades fundamentales como bañarse, vestirse, ir al baño, moverse de manera autónoma, controlar la orina y alimentarse. Tras este índice, se crearon otros para contextos más específicos como distintas enfermedades. Entre estas escalas podemos encontrar el *Índice de Barthel* [10], más amplio que el de Katz tal y como se puede ver en la 3.1, cubriendo 10 actividades y orientado a pacientes con problemas neuromusculares.

En lo que se refiere a las IADL, en 1969 *Lawson* creó un índice [11] con 8 actividades a realizar tanto dentro como fuera del hogar, entre las que se encuentran el mantenimiento del hogar, preparar la comida, lavar la ropa, o ser responsable de su propia medicación, entre otras. Originalmente, este índice presentaba un sesgo de género al no puntuar a los hombres en algunas de las actividades del hogar previamente citadas, como en la preparación de comida o el lavado de la ropa. Existen otros índices para las ADLs instrumentales como el de *Actividades de Frenchay* [12], el cual en vez de 8 presentaba hasta 15 actividades y estaba orientado a pacientes que habían sufrido un infarto.

Tabla 3.1: BADLs de los distintos índices

Actividad	Índice de Katz	Índice de Barthel
Alimentarse	✓	✓
Bañarse/Ducharse	✓	✓
Asearse	✗	✓
Vestirse/Desvestirse	✓	✓
Control de orina	✓	✓
Control de heces	✗	✓
Ir al baño	✓	✓
Moverse a una silla o cama	✓	✓
Sentarse	✗	✓
Subir y bajar escaleras	✗	✓

### 3.2. MONITORIZACIÓN INTRUSIVA VS. MONITORIZACIÓN NO INTRUSIVA

Tras haber introducido los conceptos fundamentales de las ADL y lo importantes que pueden ser para la calidad de vida de las personas mayores, es pertinente abordar ahora cómo se monitorizan estas actividades en el contexto de los hogares inteligentes. Existen diversos enfoques para esta monitorización, pero sobre todo se destaca la distinción entre estudios intrusivos y no intrusivos. Los estudios intrusivos implican la instalación de sensores específicos o dispositivos de monitorización que pueden alterar la dinámica cotidiana del hogar. Por otro lado, los estudios no intrusivos buscan minimizar la interferencia en la vida diaria del usuario al utilizar tecnologías que requieren una intervención mínima.

La percepción de lo intrusivo o no intrusivo de las tecnologías por parte de las personas mayores es un aspecto crucial en la implementación de soluciones de monitorización en sus hogares. La aceptación y comodidad de los usuarios mayores con la tecnología dependen en gran medida de factores como la transparencia en el propósito de las herramientas, la simplicidad de uso y el respeto a la privacidad [13].

En muchos casos, las personas mayores pueden intuir en lo intrusivo de una tecnología si perciben que esta interfiere significativamente con su vida diaria, requiere una instalación compleja o recopila información demasiado detallada y personal. Por otro lado, las tecnologías percibidas como no intrusivas suelen ser aquellas que se integran de manera transparente, no requieren una intervención constante y respetan la autonomía y la privacidad del usuario.

Ambos enfoques tienen sus ventajas y desafíos. La monitorización intrusiva tiende a proporcionar datos más específicos, pero puede afectar la comodidad del usuario, mientras que la no intrusiva busca equilibrar la obtención de información valiosa con el respeto a la privacidad y la comodidad de los residentes. La elección entre estos enfoques dependerá de la naturaleza específica del proyecto y de las consideraciones éticas y de privacidad que guíen la implementación.

#### 3.2.1. Monitorización no intrusiva

En el ámbito de la Vida Cotidiana Asistida por el Entorno (*Ambient Assisted Living, AAL*), los métodos de monitorización no intrusiva que se muestran en la Figura 3.1 pretenden mantener la privacidad sin dejar de proporcionar datos útiles. Los sensores ambientales instalados en el hogar pueden controlar discretamente factores ambientales como temperatura y la humedad, o incluso sonidos o ultrasonidos [14]. Los dispositivos domésticos inteligentes, como detectores de movimiento y bombillas inteligentes, proporcionan datos sobre patrones de uso, ofreciendo una visión indirecta de las rutinas diarias. Las etiquetas de identificación por radiofrecuencia (RFID) pueden colocarse en objetos de uso frecuente para proporcionar datos sobre las actividades diarias [15]. También se han creado soluciones no intrusivas utilizando sensores infrarrojos pasivos (PIR) para detectar actividades [16, 17]. La monitorización no intrusiva de la carga (NILM) destaca sobre todo por su capacidad para

controlar el uso de la energía sin necesidad de instalar equipos en cada aparato, lo que minimiza las interrupciones y favorece la escalabilidad entre diferentes.



Figura 3.1: Métodos de monitorización no intrusiva

### 3.2.2. Monitorización intrusiva

Los métodos de supervisión intrusiva, listados en la Figura 3.2 ofrecen potencialmente datos más precisos, pero plantean problemas éticos y de privacidad debido a la naturaleza íntima de la información recopilada. Algunos ejemplos son los sistemas de video [18], que, aunque proporcionan datos visuales detallados, se consideran invasivos por sus importantes problemas de privacidad. Los dispositivos que se pueden llevar puestos [19], como los teléfonos inteligentes, ofrecen información valiosa sobre las actividades diarias y los parámetros de salud, pero también pueden considerarse intrusivos. Existen estudios que utilizan combinaciones de ambas técnicas [20]. Además, la Monitorización Intrusiva de la Carga (ILM), un método menos estudiado debido a que requiere la instalación de sensores en cada aparato, también considerado perturbador por los usuarios, ofrece datos más detallados y exactitud que los enfoques NILM.

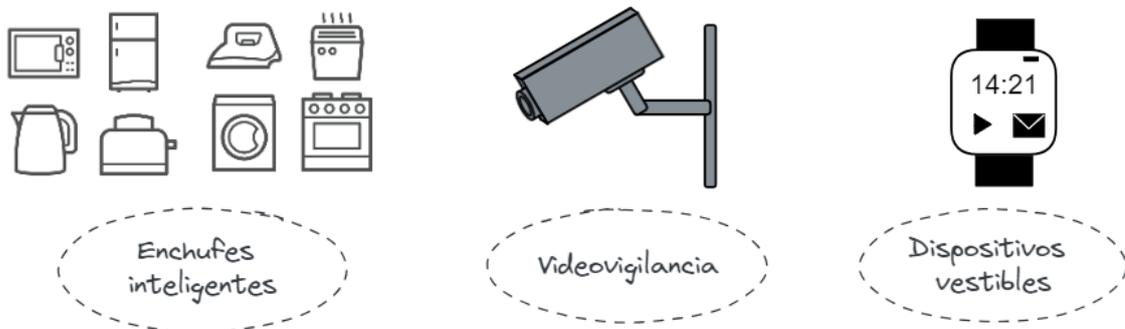


Figura 3.2: Métodos de monitorización intrusiva

### 3.3. ILM Y NILM: ENFOQUES ESPECÍFICOS EN MONITORIZACIÓN DE CARGAS ELÉCTRICAS

Dentro de este vasto panorama de monitorización, la atención se dirige especialmente hacia dos enfoques de gran relevancia: el Intrusive Load Monitoring (ILM) y el Non-Intrusive Load Monitoring (NILM). Estas metodologías se centran en la monitorización de los consumos eléctricos, una vía fundamental para comprender las actividades y comportamientos en el hogar.

**NILM (Non-Intrusive Load Monitoring):** Este enfoque se basa en analizar las firmas de carga eléctrica en todo el hogar, sin la instalación de sensores específicos en cada dispositivo. Aunque menos detallado que el ILM, su no intrusividad es evidente al no requerir modificaciones en la infraestructura eléctrica existente.

- *NILM (Non-Intrusive Load Monitoring)*: El NILM se basa en el análisis de las firmas de carga eléctrica en todo el hogar, sin la necesidad de sensores específicos en cada dispositivo [21, 22, 23]. Ofrece un enfoque no intrusivo, pero sin ofrecer nada de detalle para comprender patrones de uso general, siendo primero necesario dividir por electrodomésticos la carga y detectar tras ello las actividades [24].
- *ILM (Intrusive Load Monitoring)*: Al centrarse en el consumo eléctrico de dispositivos específicos, el ILM requiere la instalación de sensores en cada dispositivo cuyo consumo de energía se desea monitorear. A pesar de ofrecer gran detalle, se puede considerar intrusiva al ser necesaria la intervención en la infraestructura eléctrica existente en el hogar. Aun así, ofrece mayor detalle que el NILM y permite la identificación precisa de dispositivos y actividades. Gracias a la revisión de la literatura que se hizo sobre este apartado en específico, se pudo comprobar que, en comparación con el anterior, apenas existen estudios sobre él relacionados con ADLs. Es por ello que para este trabajo se utilizó este enfoque, buscando investigar otros métodos de monitorización de las ADLs distintos a los habituales.

### 3.4. DETECCIÓN DE ACTIVIDADES DE LA VIDA DIARIA

Junto a las Actividades a detectar, los tipos de monitorización y la estandarización para su detección [25], es necesario para este trabajo realizar también un estudio de las técnicas que existen para su identificación para conseguir una imagen completa del marco de trabajo a seguir. Según Mshali[26], habría que dividir estas técnicas en tres tipos:

- **Técnicas estadísticas**: Las técnicas estadísticas, especialmente los *modelos de Markov ocultos (HMM)* [27] y las *redes bayesianas* [28], son ampliamente usadas en el sector salud para modelar el comportamiento humano, incluyendo la detección y predicción de actividades. Los HMM son efectivos, pero presentan limitaciones como la dificultad para manejar grandes volúmenes de datos sensoriales y la falta de jerarquía en la representación de comportamientos. Los *campos aleatorios condicionales (CRF)* [29] se presentan como una alternativa superior en términos de clasificación y precisión, aunque requieren más procesamiento de datos. Por otro lado, las *redes bayesianas* ofrecen un marco más amplio, pero con complicaciones en la inferencia probabilística exacta. Por último, el *Modelo de Mezcla Gaussiana (GMM)* [30] ha demostrado ser útil para aprender comportamientos normales en contextos de hogares inteligentes.
- **Técnicas de inteligencia computacional**: Las técnicas de inteligencia computacional, como las *redes neuronales*, *máquinas de vectores de soporte (SVM)* [31], *minería de datos*, *árboles de decisión* y *sistemas difusos*, son utilizadas ampliamente como alternativas a los métodos estadísticos para reconocer actividades y predecir comportamientos y condiciones de salud humana. Las *redes neuronales artificiales (ANN)* [32], en particular, son herramientas matemáticas de autoaprendizaje robusto en la clasificación de datos y se han aplicado en sistemas de monitoreo de salud. Por ejemplo, se utilizan para detectar patrones anormales en signos vitales y reconocer actividades diarias, mostrando alta precisión en comparación con otros algoritmos. Estos también se han empleado ampliamente para el reconocimiento de patrones y clasificación, incluyendo la detección de caídas y el reconocimiento de actividades cotidianas. En cuanto a la minería de datos y árboles de decisión, se usan para clasificar datos de sensores y predecir comportamientos humanos. Los métodos de agrupamiento como el *K-Means* [33] y el *Fuzzy C-Means* [34] son técnicas no supervisadas también útiles para modelar comportamientos.
- **Otras técnicas**: Se pueden encontrar otras técnicas, como las impulsadas por el conocimiento, que son utilizadas para modelar el comportamiento humano, basándose en la observación de rutinas diarias y su secuencia en situaciones específicas relacionadas con tiempo, lugar, frecuencia y objetos. Estas técnicas aprovechan el conocimiento previo para desarrollar razonamiento lógico formal y deducir comportamientos, utilizándose en reconocimiento de actividades, detección de anomalías y predicción de comportamientos. Entre las técnicas más populares están los métodos basados en reglas, razonamiento basado en casos, lógica difusa y razonamiento basado

en ontologías. Los métodos basados en reglas construyen modelos descriptivos a partir de reglas primitivas, pero enfrentan limitaciones al detectar comportamientos humanos complejos directamente. La lógica difusa gestiona la incertidumbre de los datos en redes de sensores y adapta sistemas de inferencia para modelar el comportamiento. El razonamiento basado en ontologías es extensamente usado para crear sistemas conscientes del contexto en el ámbito de la salud, permitiendo el intercambio de conocimiento y un razonamiento contextual eficiente. Estos métodos se han aplicado para monitorear pacientes y mejorar la precisión en sistemas de reconocimiento de actividades domésticas.

Una vez definidas las técnicas con las cuales es posible por su naturaleza el poder monitorizar, se dividen según los dos términos ya utilizados previamente en este trabajo, *aprendizaje supervisado* y *no supervisado*, y se realiza una revisión sistemática de soluciones ya creadas con estas técnicas para distintos tipos de monitorización:

- **Algoritmos de aprendizaje no supervisado:** En el campo de la inteligencia computacional, técnicas como la minería de datos y métodos de agrupamiento ofrecen alternativas valiosas para el modelado de comportamientos sin la necesidad de datos etiquetados. El Modelo de Mezcla Gaussiana (GMM) [35] se destaca por su habilidad para aprender distribuciones de comportamientos normales en entornos como hogares inteligentes. Métodos de agrupamiento como K-Means [36, 37] y Fuzzy C-Means [38] también son esenciales para organizar grandes volúmenes de datos sensoriales en clústeres basados en características similares, ayudando a modelar y entender patrones de comportamiento humano de manera eficiente y efectiva, lo cual podría ser útil en este trabajo. Además, los Modelos de Markov Ocultos (HMM) [39, 40, 41] se utilizan para identificar patrones en secuencias de datos sin necesidad de etiquetas explícitas (aunque también se podrían utilizar para aprendizaje supervisado).
- **Algoritmos de aprendizaje supervisado:** Las técnicas de inteligencia computacional y estadísticas ofrecen poderosos algoritmos de aprendizaje supervisado para modelar el comportamiento humano y predecir actividades. Campos Aleatorios Condicionales (CRF) [42, 43, 44] son utilizados para entender secuencias y clasificar datos con alta precisión, aunque ambos enfrentan desafíos en manejo de datos y requerimientos computacionales. Las Redes Bayesianas [45] proporcionan un enfoque probabilístico complejo pero efectivo para la inferencia. Además, las Redes Neuronales Artificiales (ANN) [46], Máquinas de Vectores de Soporte (SVM) [47], y Árboles de Decisión [48] son ampliamente aplicados en sistemas de monitoreo de salud para detectar patrones anormales en signos vitales y reconocer actividades diarias. Otro tipo de red neuronal, las LSTM (*Long-Short Term Memory*, memoria a largo y corto plazo), son particularmente adecuadas para este tipo de datos, ya que tienen la capacidad de aprender y recordar patrones a lo largo del tiempo [49], y por eso podrían ser muy útiles para este proyecto. Estos modelos aprenden de datos etiquetados para realizar tareas de reconocimiento de actividades y predicción con notable precisión.

### 3.5. DETECCIÓN DE ANOMALÍAS

Aparte de la detección de actividades, otra parte fundamental en este proyecto es poder detectar aquellos comportamientos anómalos que puedan resultar significativo tras la detección de las propias ADL. Por ello, se debe realizar una revisión de los métodos para identificar anomalías en ADLs, teniendo en cuenta el modelo de sensorización a emplear en este proyecto [50], y el carácter de serie temporal del mismo. Tras esta revisión, en la que primero se buscó una definición general de la detección de anomalías [51] y de aquellas técnicas que podrían ser útiles, para después listar aquellas que en concreto que resultan adecuadas en el contexto de la actividad humana [52, 53, 54, 55]. Por tanto, se listan los siguientes métodos:

- *Técnicas de inteligencia computacional:* Como ya se ha mencionado antes, estas técnicas utilizan modelos computacionales para procesar y aprender de los datos automáticamente, permitiendo detectar patrones y anomalías, y realizar predicciones precisas. Se pueden listar los siguientes

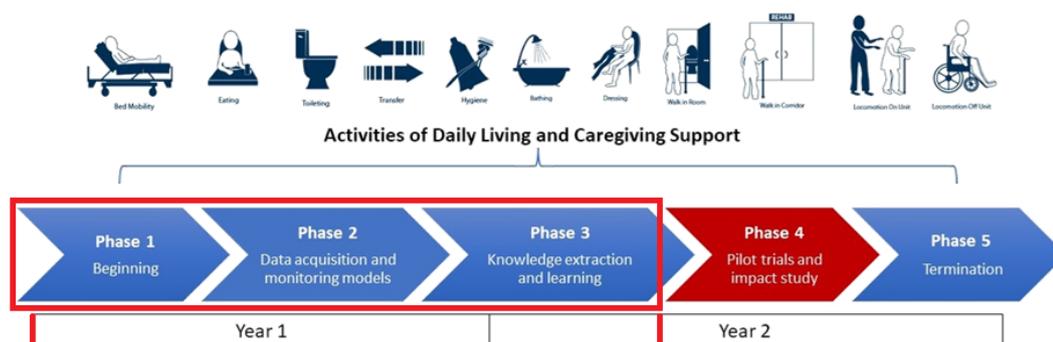
métodos:

1. **One-Class SVM:** Método de aprendizaje automático que utiliza un hiperplano para separar clases en el espacio, diseñado específicamente para distinguir entre una clase normal y anomalías. A diferencia del SVM tradicional, que se enfoca en maximizar la distancia entre dos clases, One-Class SVM concentra en delimitar una sola clase de datos normales, clasificando nuevos puntos según caigan dentro o fuera de esta región delimitada. [56, 57, 58].
  2. **Isolation Forest:** Los bosques de aislamiento son un método eficaz para detectar anomalías, ya que opera bajo el principio de que los valores atípicos son más fáciles de aislar que los valores normales. Utiliza árboles de decisión para aislar instancias, y las anomalías son identificadas como aquellas instancias que requieren menos divisiones para ser aisladas. Este método es único en el sentido de que no depende de la distancia o la densidad de los datos [59, 60].
  3. **LOF:** El factor de anomalías locales (*Local Outlier Factor*) es un enfoque no supervisado para la detección de valores atípicos que se basa en la estimación de la densidad de un punto de datos en relación con sus vecinos más cercanos. Este enfoque asume que la densidad alrededor de un valor atípico es significativamente diferente de la densidad alrededor de sus vecinos. Por lo tanto, los puntos de datos que presentan una densidad relativamente baja comparada con la de sus vecinos se consideran valores atípicos [61].
  4. **RNN:** Las Redes Neuronales Recurrentes son un tipo de redes neuronales diseñadas para procesar secuencias de datos, aprovechando su capacidad para mantener información en el 'tiempo' o a través de secuencias. Esto las hace ideales para tareas donde el contexto y el orden de los datos importan, como el análisis de series temporales o el procesamiento de lenguaje natural. En la detección de anomalías, las RNN pueden identificar comportamientos inusuales en secuencias de datos, como patrones irregulares en las actividades diarias, detectando desviaciones respecto a lo que se considera normal basado en los datos históricos aprendidos [62].
- **Técnicas estadísticas:** Las técnicas estadísticas aplican métodos matemáticos para analizar datos y extraer conclusiones sobre su comportamiento [63]. Se listan las siguientes:
    1. **Estimador de Covarianza Robusta:** La estimación habitual de covarianza de máxima verosimilitud es muy sensible a la presencia de valores atípicos en el conjunto de datos. En tal caso, este modelo más estadístico indica que sería mejor utilizar un estimador robusto de covarianza para garantizar que la estimación sea resistente a observaciones "erróneas" en el conjunto de datos [64].
    2. **CUSUM:** El algoritmo CUSUM (gráfico de sumas acumuladas o *cumulative sum control chart*) fue desarrollado para estimar el punto de cambio en el tiempo cuando un proceso que sigue una función de densidad de probabilidad conocida cambia repentinamente el valor de uno de sus parámetros. Este algoritmo es útil, por ejemplo, para detectar cambios abruptos en patrones temporales, como un cambio en el horario de despertar de una persona debido a una enfermedad [65, 66].

# Metodología

En este capítulo se detalla la metodología empleada para la planificación y desarrollo del trabajo, explicando por qué se ha elegido y cómo se ha aplicado dicha metodología. Después, se exponen los recursos *hardware* y *software* que son necesarios para desarrollar el proyecto.

En este proyecto estatal, dividido en cinco fases, tal y como se puede ver en la Fig. 4.1, se han abordado partes de las tres primeras etapas durante el primer año y medio: la fase inicial, centrada en la selección de la tecnología adecuada (no multimodal de momento) y la definición de los objetivos del proyecto, seguida de la fase de recogida de información y diseño de modelos de monitorización, donde se establecieron los métodos para extraer datos de los dispositivos seleccionados, y la tercera fase, donde se desarrollaron modelos analíticos para extraer conocimiento de la información monitoreada. Esto permitirá la prueba de soluciones y el diseño de capas proactivas de interacción en las siguientes fases del proyecto.



**Figura 4.1:** Fases del proyecto

Cada una de las fases del proyecto está destinada a abordar diferentes aspectos del objetivo del trabajo, desde la selección inicial de la tecnología hasta el desarrollo de un marco integral para la recopilación y análisis de datos. En el corazón de este enfoque se encuentra la metodología de prototipado rápido (PR)[67], una estrategia que prioriza la iteración y la adaptabilidad, permitiendo responder dinámicamente a los retos y a la retroalimentación en tiempo real.

La primera de las fases completadas este año se centró en la selección y evaluación de los enchufes inteligentes disponibles en el mercado, buscando dispositivos que no solo cumplieran con los requisitos técnicos esenciales, sino que también ofrecieran la fiabilidad y la facilidad de uso necesarias para una integración en los hogares de los participantes. Esta etapa inicial fue fundamental para establecer una base sobre la cual construir el resto del proyecto, asegurando que las herramientas seleccionadas fueran capaces de proporcionar los datos necesarios para el análisis posterior.

La segunda fase se dedicó a desarrollar métodos eficientes para extraer información de los enchufes inteligentes elegidos. Esta etapa involucró tanto la configuración técnica como la programación necesaria para facilitar la recopilación de datos, presentando desafíos como la integración de tecnologías y seguridad de datos. A través de un proceso iterativo de prueba y mejora, se fue capaz de

establecer un flujo de datos confiable, abriendo el camino para análisis más profundos y la generación de resultados significativos sobre las actividades en la vida doméstica.

Antes del inicio del primer prototipo en marzo, en enero y febrero se realizó una planificación y definición de requisitos. Esta etapa preliminar fue importante para establecer una base para el proyecto, permitiendo identificar y detallar los requisitos específicos que cada uno de los cinco prototipos debería cumplir a lo largo del año y medio. Este período de planificación aseguró una hoja de ruta y objetivos bien definidos para guiar el desarrollo de cada prototipo, facilitando así un proceso más estructurado y enfocado para el proyecto en su conjunto. El resto del año se configuró como se puede apreciar en la Fig. 4.2. Cabe destacar que el mes de agosto fue en casi su totalidad vacacional, y que desde septiembre las horas de trabajo se doblaron porque hasta entonces se trabajaba a tiempo parcial, compaginándolo con los estudios del máster:

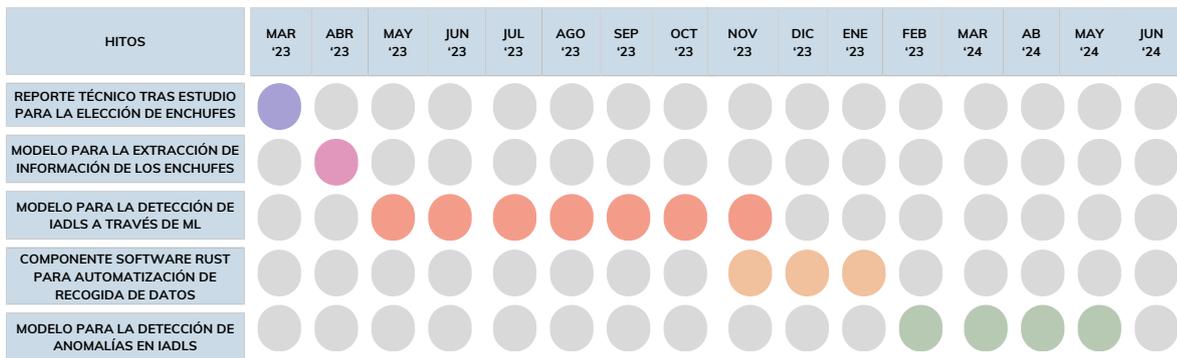


Figura 4.2: Diagrama de Gantt del proyecto

### 4.1. METODOLOGÍA DE DESARROLLO SOFTWARE

El ciclo de vida del desarrollo del software (SDLC, *Software Development Life-Cycle*) llevado a cabo en el proyecto (Fig. 4.3) es la metodología de prototipado rápido, la cual tiene un enfoque dinámico y flexible en el desarrollo de proyectos, particularmente efectivo en campos como el desarrollo de software, ingeniería de sistemas, y diseño de productos. Esta metodología se centra en la creación de prototipos funcionales en fases tempranas del desarrollo para explorar, iterar y validar ideas y requisitos de forma práctica y colaborativa. A diferencia de los enfoques tradicionales que requieren una definición completa de los requisitos antes de cualquier desarrollo sustancial, el prototipado rápido invita a una experimentación continua y a la adaptación, permitiendo ajustes basados en la retroalimentación real de las partes interesadas.

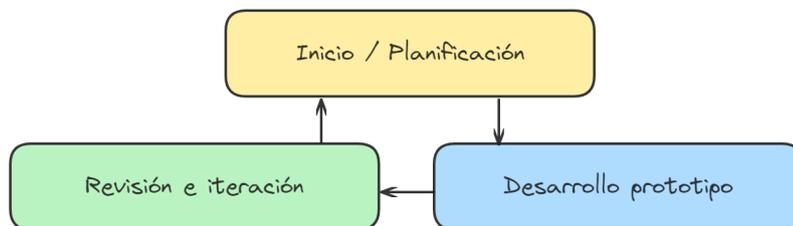


Figura 4.3: Metodología PR aplicada al proyecto

En el contexto de este proyecto, aplicar esta metodología significa avanzar a través de una serie de prototipos, cada uno diseñado para probar y mejorar aspectos específicos del sistema. A continuación, se describe cómo se aplica la metodología de prototipado rápido a cada uno de los cinco prototipos del proyecto:

## 4.2. PRIMER PROTOTIPO: REPORTE TÉCNICO TRAS ESTUDIO PARA LA ELECCIÓN DE ENCHUFES INTELIGENTES

### 4.2.1. Inicio y planificación:

El proyecto comienza con la identificación de la necesidad de monitorear las actividades de la vida diaria a través de enchufes inteligentes. Se establecen los criterios para la selección de estos dispositivos, considerando factores como la conectividad, la seguridad, la compatibilidad con diferentes lenguajes y la facilidad para extraer datos.

### 4.2.2. Desarrollo del prototipo:

Se realiza un estudio comparativo de los enchufes inteligentes disponibles en el mercado, utilizando conocimiento previo de otros proyectos realizados con enchufes inteligentes dentro del laboratorio **MAmI** (*Modelling Ambient Intelligence*), y teniendo en cuenta la cantidad de datos que proporciona cada enchufe.

### 4.2.3. Revisión e iteración:

Basándose en la retroalimentación y los resultados obtenidos, se selecciona el enchufe inteligente más adecuado. Se descartó añadir otra fuente de información como bombillas inteligentes o sensores de movimiento para centrarse en la monitorización de consumo energético de distintos electrodomésticos.

## 4.3. SEGUNDO PROTOTIPO: MODELO PARA LA EXTRACCIÓN DE INFORMACIÓN DE LOS ENCHUFES

### 4.3.1. Inicio y planificación:

Con los enchufes seleccionados, se define la planificación técnica para la extracción eficaz de los datos, incluyendo la frecuencia de recolección, los protocolos de comunicación y la información que se considerará relevante para la recolección, evitando así recoger aquella información que no aporte ningún conocimiento.

### 4.3.2. Desarrollo del prototipo:

Se implementa una solución técnica para conectar con los enchufes a través de la aplicación móvil de dichos dispositivos, para poder después extraer los datos necesarios. Esto puede involucrar el desarrollo de una API personalizada, un escaneo de la red o la creación de una interfaz para visualizar la información.

### 4.3.3. Revisión e iteración:

La solución se prueba en un entorno real, una vez adquiridos los dispositivos, recopilando datos y ajustando el sistema en función de la calidad de los datos recogidos.

## 4.4. TERCER PROTOTIPO: MODELO PARA LA DETECCIÓN DE IADLS A TRAVÉS DE MACHINE LEARNING

### 4.4.1. Inicio y planificación:

La meta inicial definida para el modelo de inteligencia artificial será el poder identificar actividades cotidianas mediante la información recolectada. Esto permitirá reconocer patrones y desviaciones en los datos obtenidos de los enchufes inteligentes.

Se definen los objetivos específicos del modelo de IA, como identificar actividades de la vida diaria a través de la información recogida que permitan reconocer patrones y desviaciones que correlacionen los datos de los enchufes con esas actividades específicas.

#### 4.4.2. Desarrollo del prototipo:

Utilizando los datos recogidos, se desarrolla un modelo preliminar basado en técnicas de IA, aplicando técnicas de aprendizaje supervisado y no supervisado. Se realizan pruebas iniciales para evaluar su capacidad para identificar actividades con las distintas técnicas.

#### 4.4.3. Revisión e iteración:

El modelo se ajusta y perfecciona basándose en su precisión y eficacia, iterando a través de diferentes configuraciones, algoritmos o enfoques de preprocesamiento de datos.

### 4.5. CUARTO PROTOTIPO: COMPONENTE SOFTWARE RUST PARA LA RECOGIDA AUTOMÁTICA DE INFORMACIÓN

#### 4.5.1. Inicio y planificación:

Se define la necesidad de un código robusto para automatizar la recolección de datos. Rust se elige por su rendimiento y seguridad. Se especifican las funcionalidades que debe cumplir la API, eligiendo también la capa de almacenamiento a la que se destinarán los datos.

#### 4.5.2. Desarrollo del prototipo:

Se desarrolla una versión inicial del código, implementando la lógica para interactuar con los enchufes inteligentes y recolectar y almacenar datos de forma automatizada y segura.

#### 4.5.3. Revisión e iteración:

Se prueba el código en condiciones reales en un ambiente controlado, ajustándose para mejorar su eficiencia, facilidad de uso y compatibilidad con sistemas existentes o futuros.

### 4.6. QUINTO PROTOTIPO: MODELO PARA LA DETECCIÓN DE ANOMALÍAS EN IADLS

#### 4.6.1. Inicio y planificación:

Se realiza una revisión exhaustiva del estado del arte para establecer cómo se ha abordado previamente la detección de anomalías en las Actividades Instrumentales de la Vida Diaria (IADLs). Se definen los objetivos específicos a cumplir por el modelo.

#### 4.6.2. Desarrollo del prototipo:

Con base en los datos previamente analizados y las actividades extraídas, se implementan funciones necesarias para desarrollar un modelo capaz de detectar anomalías. Se aplicarán técnicas estadísticas para realizar la detección de desviaciones en las IADLs.

#### 4.6.3. Revisión e iteración:

El modelo se ajusta y perfecciona basándose en su precisión y eficacia en las pruebas a través de diferentes configuraciones. Estas pruebas se llevan a cabo con los datos disponibles para asegurar el correcto funcionamiento del código y la eficacia en la detección de anomalías.

Como se puede ver en la Fig. 4.4, a través de este enfoque iterativo e incremental, cada prototipo construye sobre los aprendizajes del anterior, permitiendo una adaptación continua a nuevos desafíos y requisitos. Este proceso no solo facilita el manejo de la complejidad del proyecto, sino que también asegura que el producto final esté alineado con las necesidades reales de los usuarios, maximizando así su valor y efectividad.

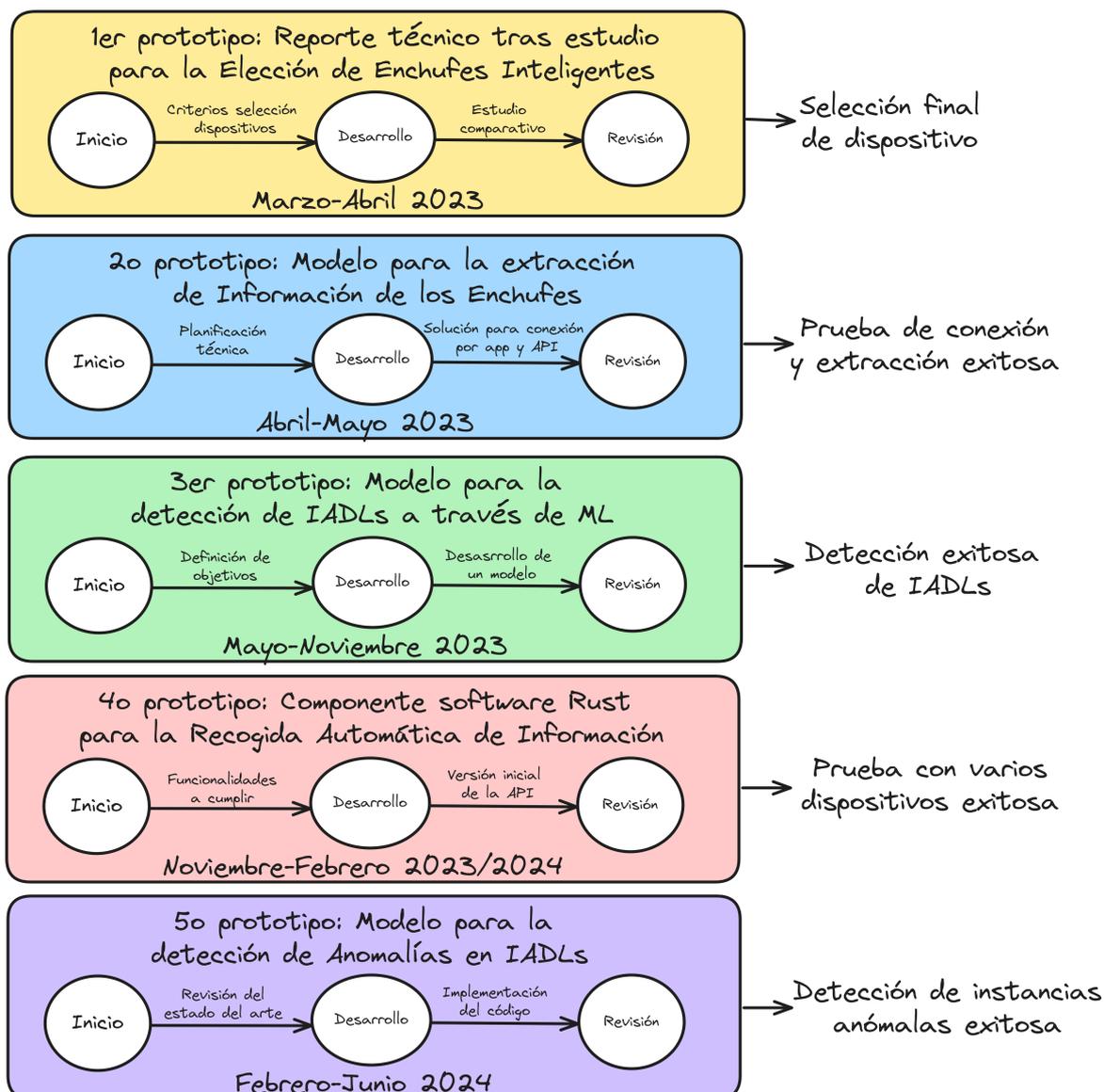


Figura 4.4: Resumen de los prototipos y sus resultados

## 4.7. RECURSOS PARA EL DESARROLLO

### 4.7.1. Herramientas software y lenguajes de programación

1. *Microsoft Teams*: Herramienta de comunicación online que permite crear reuniones, llamar y colaborar con otros usuarios. Todas las reuniones no presenciales llevadas a cabo a lo largo del proyecto se realizaron en esta aplicación. <sup>1</sup>
2. *Dbdiagram*: Herramienta simple y gratis para desarrollar diagramas entidad-relación a través de código. Usado para diseñar el diagrama de la base de datos del proyecto. <sup>2</sup>
3. *Firebase*: Plataforma de desarrollo de aplicaciones web que ayuda a compilar y desarrollar estas apps, contando con el respaldo de Google. Permite el desarrollo para varias plataformas y cuenta con numerosas herramientas de almacenamiento. En este proyecto, solo se utilizará este almacenamiento NoSQL que permiten sus bases de datos en tiempo real para realizar pruebas de almacenamiento en las primeras etapas del proyecto. <sup>3</sup>

<sup>1</sup><https://www.microsoft.com/es-es/microsoft-teams/log-in>

<sup>2</sup><https://dbdiagram.io/home/>

<sup>3</sup><https://firebase.google.com/>

4. *MongoDB Atlas*: MongoDB Atlas es un conjunto integrado de servicios de datos centrados en una base de datos en la nube diseñada para acelerar y simplificar la forma de construir con datos. En este proyecto servirá como la base de datos principal del proyecto. <sup>4</sup>
5. *Rust*: Rust es un lenguaje de programación que ofrece rendimiento, fiabilidad y productividad para aplicaciones de software. Rust es increíblemente rápido y eficiente con la memoria. El rico sistema de tipos de Rust y su modelo de propiedad (*ownership*) garantizan seguridad de memoria y seguridad en hilos, y te permiten eliminar muchas clases de bugs, reportándose a la hora de compilar. Se utilizó para realizar la API que recoge información de los enchufes. <sup>5</sup>
6. *Python*: Python es un lenguaje de alto nivel de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código, se utiliza para desarrollar aplicaciones de todo tipo. Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, dinámico y multiplataforma. Utilizado para desarrollar los modelos que extraen información de los datos recopilados. <sup>6</sup>
7. *Google Colab*: Colab es un servicio alojado de Jupyter Notebook que no requiere configuración y proporciona acceso gratuito a recursos informáticos, incluidas GPU y TPU. Colab es especialmente adecuado para el aprendizaje automático, la ciencia de datos y la educación. Por tanto, se utilizaron estos recursos para desarrollar en este entorno los modelos. <sup>7</sup>
8. *Visual Studio Code*: Editor de código eficaz y ligero con herramientas integradas para implementar código fácilmente en cualquier sistema operativo, permitiendo adaptarse a casi cualquier lenguaje de programación con sus extensiones. Se usó para el desarrollo de la API. <sup>8</sup>
9. *GitHub*: Sitio web y servicio en la nube que permite almacenar y administrar código, al igual que llevar un registro y control de cualquier cambio sobre este código a través de Git, un sistema de control específico de versiones. Se almacena todo el código del proyecto en este servicio en la nube. <sup>9</sup>
10. *Overleaf*: Overleaf es una plataforma web que te permite crear, editar y compartir documentos LaTeX desde cualquier lugar. La documentación del proyecto se creó utilizando esta herramienta. <sup>10</sup>

#### 4.7.2. Herramientas Hardware

Para todo el desarrollo del proyecto se utilizó el mismo ordenador portátil desde el que se realizaba toda la codificación. Uno de los aspectos más importantes del proyecto serán los enchufes seleccionados tras la primera iteración, los cuales se especifican a continuación. Tras la tercera iteración, se probó el modelo de inteligencia artificial en un ordenador de sobremesa con una GPU lo suficientemente potente del laboratorio MAmI. Aparte, para probar físicamente el desarrollo de la API tras la cuarta iteración se usaron dos modelos de Raspberry Pi viejos que sirvieran como puente entre los enchufes y la red, y se compraron 5 modelos de las últimas Raspberry Pi, la 5, con el objetivo de un mejor rendimiento a la hora de posibles pruebas piloto.

##### 1. Ordenador portátil.

- **Modelo:** Acer Aspire E5-575G-57RL
- **CPU:** Intel(R) Core(TM) i5-7200U CPU @ 2.50 GHz 2.70 GHz.
- **RAM:** 8,00 GB (7,87 GB usable)

---

<sup>4</sup><https://www.mongodb.com/atlas/database>

<sup>5</sup><https://www.rust-lang.org/es>

<sup>6</sup><https://www.python.org/>

<sup>7</sup><https://colab.research.google.com/>

<sup>8</sup><https://code.visualstudio.com/>

<sup>9</sup><https://github.com/>

<sup>10</sup><https://www.overleaf.com/>

- **GPU:** NVIDIA GeForce GTX 950M
- **Sistema Operativo:** Windows 10
- **Pantalla:** 15.60 Pulgadas, 16:9, 1366 x 768 px

## 2. Enchufes Tapo.

- **Modelo:** Tapo P110
- **Conectividad:** IEEE 802.11b/g/n
- **Tipo de Wi-Fi:** 2,4 GHz
- **GPU:** NVIDIA GeForce GTX 950M
- **Requerimientos de energía:** 220-240 V AC 50/60 Hz.
- **Carga máxima:** 3680 W, 16A



Figura 4.5: Enchufe Tapo P110 <sup>11</sup>

## 3. Ordenador de sobremesa.

- **CPU:** Intel Core i7-10700F CPU @ 2.90GHz
- **RAM:** 32,00 GB (31,9 GB usable)
- **GPU:** NVIDIA GeForce RTX 3060 TI
- **Sistema Operativo:** Windows 10

## 4. Raspberry Pi Modelo 3.

- **CPU:** ARM Cortex A53 Quad-Core 1.2 GHz
- **RAM:** 1 GB LPDDR2 SDRAM
- **Conectividad:** IEEE 802.11b/g/n, Bluetooth, LAN
- **Puertos:** RJ45, HDMI, 4 USB, puerto de visualización MIPI DSI, puerto de cámara MIPI CSI, SD Socket y conector de alimentación 2.5 A

## 5. Raspberry Pi Modelo 3B+.

- **CPU:** Broadcom BCM2837B0, Cortex-A53 SoC de 64 bits a 1,4 GHz
- **RAM:** 1 GB LPDDR2 SDRAM
- **Conectividad:** IEEE 802.11b/g/n, Bluetooth, LAN y Gigabit Ethernet
- **Puertos:** RJ45, HDMI, 4 USB, puerto de visualización MIPI DSI, puerto de cámara MIPI CSI, SD Socket y conector de alimentación 5 / 2.5 A

## 6. Raspberry Pi Modelo 5.

- **CPU:** CPU Arm Cortex-A2 de cuatro núcleos y 4 bits a 64,76 GHz
- **GPU:** VideoCore VII, compatible con OpenGL ES 3.1, Vulkan 1.2
- **RAM:** 4 GB de SDRAM LPDDR4X-4267.
- **Conectividad:** Wi-Fi 5, Gigabit Ethernet y Bluetooth 5.0 / BLE
- **Puertos:** 2 micro HDMI (hasta 2 x 4K 60Hz simultáneas), 2 USB 3.0, 2 USB 2.0, 1 Gigabit Ethernet con PoE opcional, 2 MIPI de 4 pistas, 1 PCIe 2.0 x1 y 1 GPIO de 40 pines
- **Almacenamiento:** SSD M.2 2280 Kioxia Exceria NVMe SSD 500GB



Figura 4.6: Raspberry Pi 5 en su carcasa

---

## CAPÍTULO 5

# Resultados

---

En esta sección se describirá la aplicación del método de trabajo presentado en el capítulo 4.

### 5.1. PLANIFICACIÓN PREVIA

Antes del comienzo del primer prototipo, varias reuniones se llevaron a cabo para definir y concretar qué actividades era necesario que se desarrollaran dentro del mismo. En los dos primeros meses de 2023 se expuso en qué consistía el proyecto **TAICare** del cual nace este TFM. El objetivo principal del proyecto era apoyar la prestación de cuidados a personas mayores que viven solas en su domicilio para mejorar su calidad de vida y la de sus cuidadores, haciendo uso de tecnologías innovadoras, sostenibles y procedimientos inteligentes. Por tanto, las actividades que se esperaban llevar a cabo en este año y media eran:

1. Diseño de un *framework* para la adquisición de datos de monitorización en el hogar y su relación con las ADLs de personas mayores.
2. Diseño de la configuración para la autenticación o registro de dispositivos de forma escalada, y decisiones sobre adquisición de datos y almacenamiento.
3. Diseño y adaptación de dispositivos para su integración en el sistema (incluyendo estudio del nuevo estándar **Matter** [68]).
4. Implementación de servicios para la monitorización a bajo nivel. Realización de simulaciones y uso de dispositivos reales (*smartplugs*, *smartbulbs*, etc.)
5. Detección de anomalías en los datos recogidos y procesados tras ser adquiridos en los pasos anteriores.

Para cada uno de los prototipos explicados en el capítulo anterior, se diseñó una iteración que lo envolviera:

### 5.2. ITERACIÓN 1: REPORTE TÉCNICO TRAS ESTUDIO PARA LA ELECCIÓN DE ENCHUFES INTELIGENTES

Tras una exhaustiva revisión del estado del arte sobre la monitorización de actividades de la vida diaria en hogares inteligentes, se tomó la decisión de utilizar enchufes inteligentes como la herramienta principal para el seguimiento con el objetivo de crear una solución innovadora, debido a que otras soluciones habían sido más estudiadas. Esta elección se fundamenta en la capacidad de los enchufes inteligentes para ofrecer un método menos intrusivo de monitoreo de carga (ILM), permitiendo una recopilación de datos detallada y precisa sin interferir significativamente en la vida diaria de los usuarios. Este enfoque se alinea con la tendencia actual hacia soluciones más discretas y respetuosas con la privacidad en el ámbito de los hogares inteligentes. Al elegir el enchufe inteligente para el proyecto, se consideraron varios aspectos clave:

1. **Conectividad y Compatibilidad:** La capacidad del enchufe para integrarse sin problemas con diferentes sistemas y estándares de hogares inteligentes. Aquí entra en juego lo comentado

previamente sobre el estándar **Matter**, ya que se debía buscar un enchufe que diera soporte en ese momento o estuviera planeado que lo diera en un futuro próximo.

2. **Fiabilidad:** La consistencia en el rendimiento y la mínima necesidad de mantenimiento o reinicio. Es decir, un enchufe que se conecte de manera segura y pueda mantener la conexión sin necesidad de interferencia del usuario o cuidador.
3. **Facilidad de Uso:** Una instalación intuitiva tanto para el uso como para la gestión diaria.
4. **Precisión en el Monitoreo:** La habilidad para recoger datos precisos y detallados sobre el consumo de energía, con un nivel de granularidad aceptable.
5. **Seguridad:** Protocolos robustos de seguridad para proteger los datos de los usuarios.

Entre los candidatos evaluados, aparte del elegido, algunos destacaron por sus características particulares:

- **TP-Link HS110:** Ofrece un monitoreo de energía detallado y es compatible con una amplia gama de dispositivos, pero su tamaño más grande puede ser un inconveniente. Esta opción fue considerada, ya que este modelo permite una simulación de sus datos como primera aproximación al monitoreo de energía. Se descartó por llevar mucho tiempo en el mercado y considerarse inseguro al no estar actualizado.

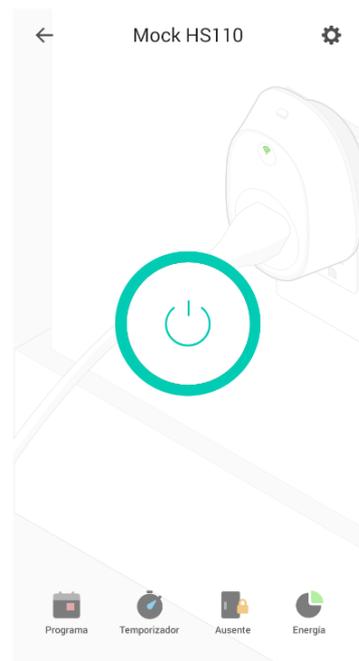


Figura 5.1: Simulación de HS110

- **Xiaomi Smart Plug 2:** Precio similar al resto, ofrece la opción de poder conectarse a él a través de Bluetooth, aparte de por la red.
- **Amazon Smart Plug:** Se integra perfectamente con el ecosistema Alexa, pero su funcionalidad de monitoreo de energía es más limitada.
- **TP-Link Tapo P110:** Evolución del HS110, el cual es una versión más obsoleta. También ofrece un monitoreo de energía detallado y es compatible con una amplia gama de dispositivos y tecnologías, e incluso tiene intención de integrar el estándar Matter en futuras actualizaciones del enchufe.<sup>1</sup>

Tras evaluar todas las opciones, el enchufe elegido fue el **Tapo P110** de TP-Link, que no solo cumple con todos los criterios mencionados, sino que también ofrece un equilibrio óptimo entre

<sup>1</sup><https://www.tapo.com/es/product/smart-plug/tapo-p110m/>

funcionalidad, precio y fiabilidad. Además, el compromiso de TP-Link con la innovación y la compatibilidad futura, como lo demuestra su intención de integrar Matter asegura que el Tapo P110 sea una elección a largo plazo que se alinea con los requisitos del proyecto de adoptar tecnologías abiertas y unificadas para la conectividad en hogares inteligentes.

### 5.3. ITERACIÓN 2: MODELO PARA LA EXTRACCIÓN DE INFORMACIÓN DE LOS ENCHUFES

Para maximizar la eficiencia y efectividad en la extracción y clasificación de información de los enchufes inteligentes, se realizó una evaluación de las distintas plataformas de extracción de datos disponibles. Esta evaluación estratégica es crucial para seleccionar la opción que mejor se integre con la arquitectura diseñada.

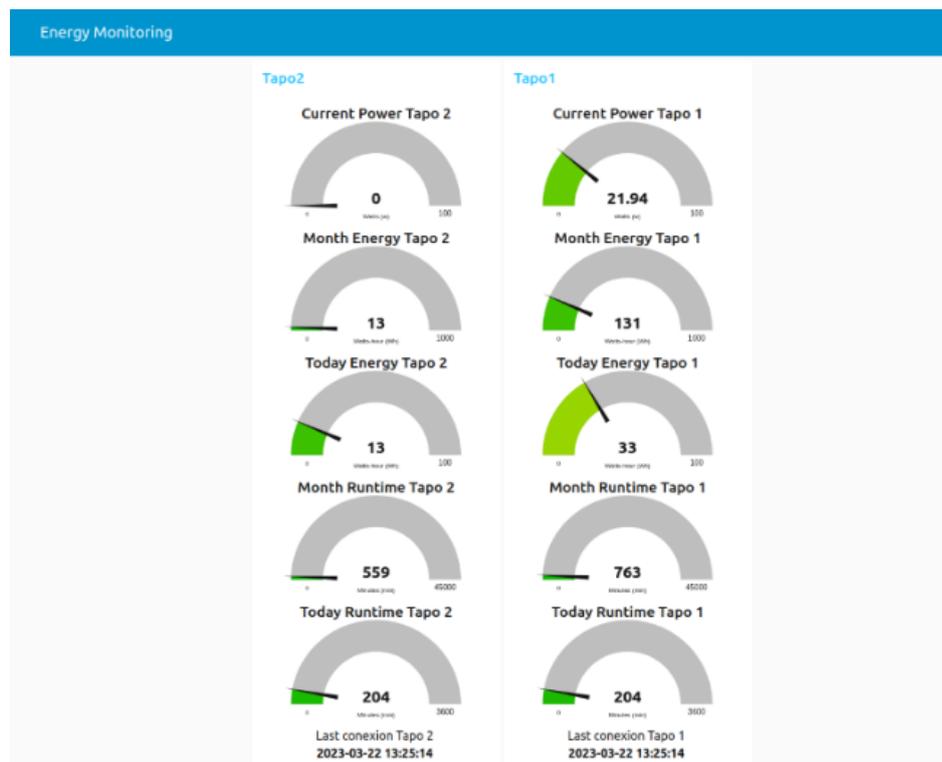


Figura 5.2: Información en Node-RED

Entre las plataformas consideradas está Node-RED, una herramienta de programación visual basada en Node.js, que se destaca por su interfaz intuitiva y una extensa biblioteca de nodos. Node-RED facilita la integración y automatización de dispositivos IoT como sensores o enchufes inteligentes, lo que la convierte en una candidata ideal para manejar datos de enchufes inteligentes dentro de nuestra infraestructura, explicada también en la asignatura de Sistemas Ciberfísicos del Máster de Ingeniería, facilitará la gestión de los flujos de datos entre los enchufes inteligentes y los sistemas de análisis. En la Fig. 5.2 se puede ver un ejemplo de cómo funciona esta plataforma, mostrando información en tiempo real sacada de los enchufes, como la potencia que está pasando por un enchufe, o el tiempo diario o mensual que ha estado conectado el enchufe.

Tras realizar estas pruebas y cerciorar la viabilidad de la extracción de información a través de los enchufes, se pasó a la definición de un *framework* para la adquisición y procesado de información.

El *framework* se articula en torno a una infraestructura bien definida, visible en el diagrama de secuencia de la Figura 5.3 que comienza en la capa de bajo nivel con los enchufes inteligentes, que constituyen los dispositivos iniciales de recolección de datos. Estos enchufes capturarán como ya se ha comentado información detallada sobre el uso de energía, funcionando como los sensores primarios en nuestro ecosistema de monitoreo.

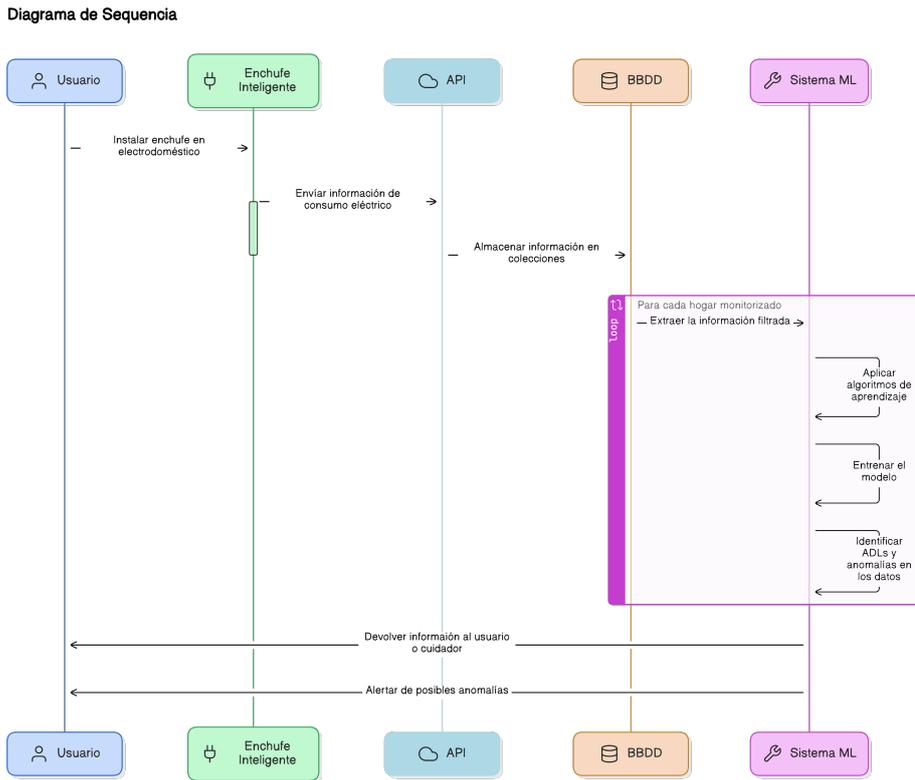


Figura 5.3: Diagrama de secuencia del trabajo de investigación

Desde los enchufes inteligentes, la información recopilada se transmite a los *gateways* o pasarelas. Estos dispositivos actúan como intermediarios, facilitando la comunicación entre los enchufes y la infraestructura de procesamiento de datos. La interacción entre los *gateways* y los enchufes inteligentes se lleva a cabo mediante peticiones REST, las cuales se autentican utilizando *tokens* obtenidos a través de credenciales de usuario y contraseña. Estas credenciales corresponden a las utilizadas durante el registro de los enchufes, asegurando así una comunicación segura.

Una vez que los datos son recopilados por los *gateways*, el siguiente paso en la arquitectura implica su almacenamiento en una base de datos centralizada. La elección de la base de datos es crucial, ya que debe ser capaz de manejar el volumen y la complejidad de los datos recolectados, además de garantizar la seguridad y la integridad de la información. Esta base de datos sirve como el repositorio principal desde donde se accede a los datos para su análisis y procesamiento posterior.

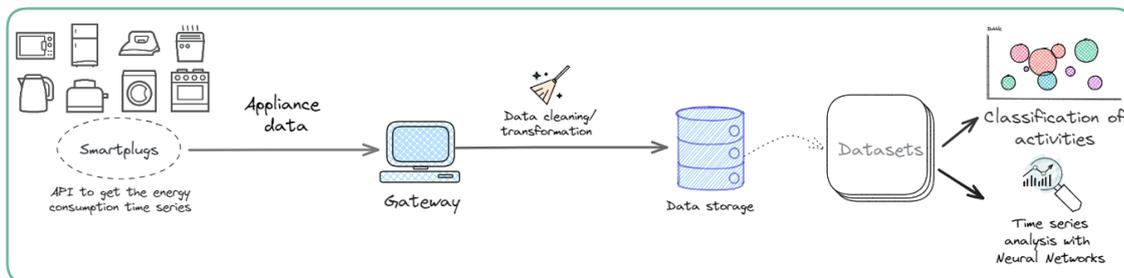


Figura 5.4: Marco de adquisición (*framework*) del proyecto

Finalmente, los datos almacenados en la base de datos se organizan y procesan en *datasets* específicos, los cuales son analizados para extraer patrones significativos y detectar actividades de la vida diaria (ADLs). Este análisis permite identificar comportamientos habituales y anomalías en las rutinas de los usuarios finales, ofreciendo información valiosa que puede utilizarse para mejorar la calidad de vida y personalizar los servicios y las intervenciones.

Para validar la viabilidad y eficacia del marco de adquisición o *framework* propuesto para el proyecto, visible en la Fig. 5.4, se llevó a cabo una prueba en local que simulaba el entorno de implementación final. En esta prueba, los datos procedentes de los enchufes inteligentes eran recopilados por un *gateway* y transmitidos utilizando el protocolo MQTT, un estándar ligero de mensajería ideal para la comunicación entre dispositivos en el Internet de las Cosas (IoT). Estos datos eran luego enviados a una base de datos InfluxDB, seleccionada por su capacidad para manejar series temporales de manera eficiente, lo que facilitaba la posterior visualización y análisis de los datos en Grafana, como se puede ver en la Fig. 5.5.

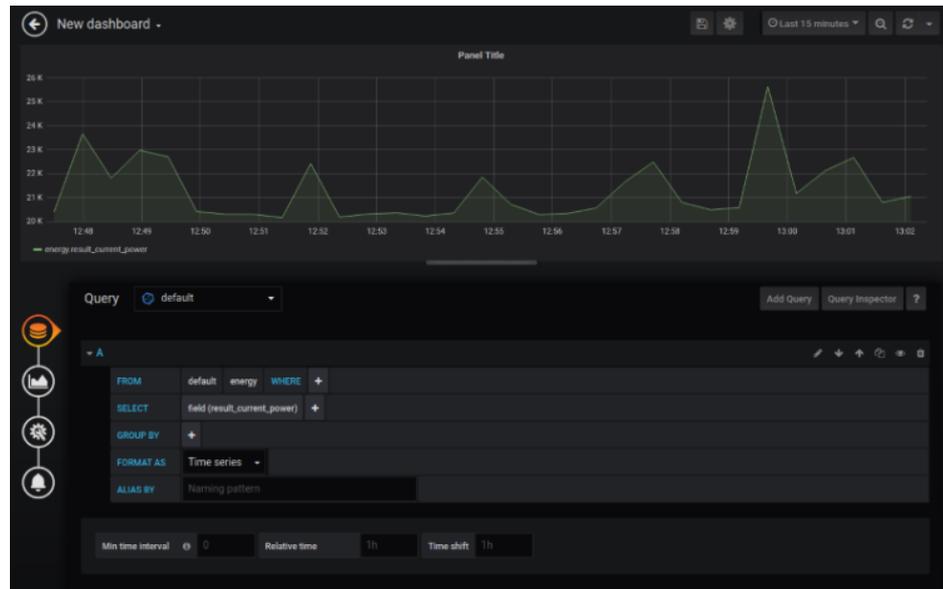


Figura 5.5: Información en Grafana

Este ejercicio preliminar fue fundamental para comprender las necesidades específicas de almacenamiento y procesamiento de datos del proyecto. A través de esta configuración de prueba, se hizo evidente que una base de datos NoSQL sería la opción más adecuada para la arquitectura final, debido a su flexibilidad, escalabilidad y capacidad para manejar grandes volúmenes de datos no estructurados, características esenciales para el análisis de datos y la detección de Actividades de la Vida Diaria (ADLs).

## 5.4. ITERACIÓN 3: MODELO PARA LA DETECCIÓN DE IADLS

### 5.4.1. Elección de *dataset* público

Dada la naturaleza de esta investigación en la monitorización de actividades de la vida diaria a través de enchufes inteligentes, al no haber llevado a cabo aun las evaluaciones reales no se dispone de información procedente en entornos reales que pueda ser utilizada para extraer conocimientos mediante técnicas de aprendizaje automático. Ante esta situación, se ha decidido realizar un estudio de *datasets* públicos que ofrecieran datos específicos sobre el monitoreo eléctrico utilizando varios enchufes inteligentes, con el objetivo de comprender mejor los patrones de consumo y las posibles aplicaciones de estos datos en nuestro proyecto.

Entre los diversos conjuntos de datos evaluados, destacaron especialmente UK-DALE (*UK Domestic Appliance-Level Electricity*)[69], ENERTALK[70] y REDD (*Reference Energy Disaggregation Data Set*)[71], reconocidos por su riqueza de información y su relevancia para estudios en el ámbito de la monitorización eléctrica en hogares. Sin embargo, tras un análisis detallado de las características como duración, hogares, la accesibilidad y la aplicabilidad de los diferentes *datasets*, se tomó la decisión de utilizar el conjunto de datos REFIT: *Electric Load Measurements* [72].

Dataset	Lugar	Duración	Casas	Electrodomésticos
REDD	EE.UU.	3 19 días	6	10 24
REFIT	RU	2 años	20	9
UK-DALE	RU	2,5 años	5	5 54
ENERTALK	Corea del Sur	29 122 days	22	1 7

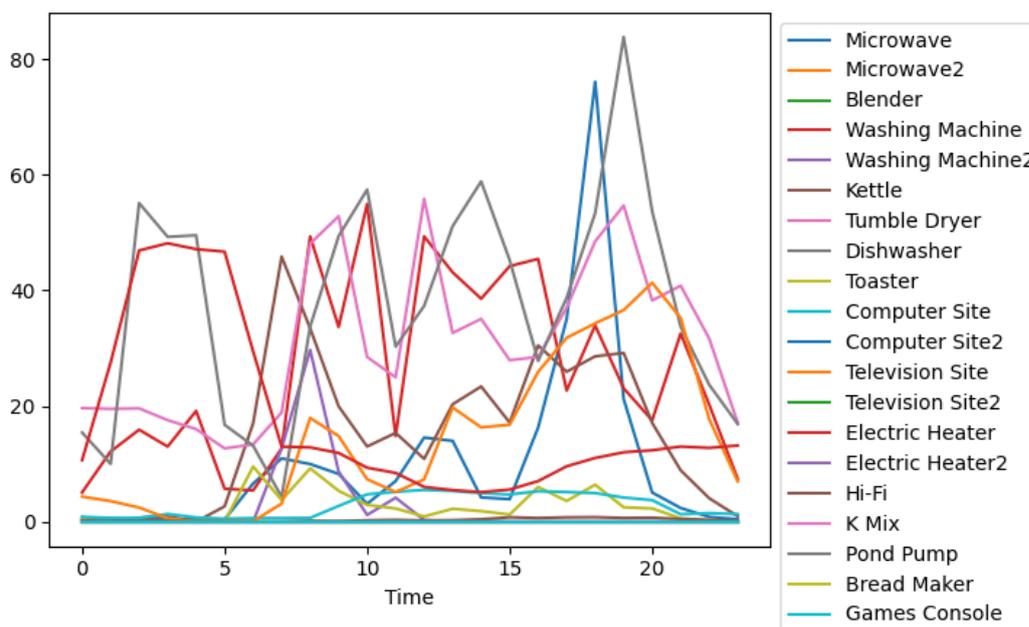
**Tabla 5.1:** Comparación de *datasets* públicos

**REFIT**, con su enfoque en las mediciones de carga eléctrica residencial en el Reino Unido, ofrece un equilibrio óptimo entre la granularidad de los datos, la variedad de los dispositivos monitorizados y la relevancia para el contexto de investigación. Este conjunto de datos no solo proporciona información detallada sobre el consumo energético a nivel de aparato en hogares reales, sino que también se alinea estrechamente con el objetivo de entender y modelar las actividades de la vida diaria a través de la monitorización del uso de energía. La elección de REFIT como la principal fuente de datos externos marca un paso crucial en la investigación, proporcionándonos una base sólida sobre la cual desarrollar en las etapas iniciales del proyecto.

#### 5.4.2. Extracción de información del *dataset*

Una vez seleccionado el *dataset* REFIT, se procedió a una etapa de preparación y limpieza de datos. El objetivo principal de este proceso fue identificar y seleccionar aquellos electrodomésticos que ofrecieran un mayor potencial para la detección de actividades cotidianas significativas, descartando aquellos dispositivos con patrones de uso menos claros o irrelevantes para nuestro estudio, como por ejemplo frigoríficos o congeladores, los cuales no cambian su potencia independientemente de la interacción del usuario. Esta selección se realizó mediante un análisis detallado de los perfiles de consumo, identificando y preservando los datos relacionados con electrodomésticos y otros aparatos electrónicos que son comúnmente utilizados en actividades diarias clave, como podemos ver en la Fig. 5.6.

Además, se limpiaron los valores que pudieran distorsionar el análisis, en este caso instancias con todos los valores a 0, asegurando así que solo los datos relevantes fueran conservados para el estudio. Este proceso de filtrado fue necesario para construir una base sobre la cual aplicar posteriormente técnicas avanzadas de aprendizaje automático, para avanzar así en el trabajo de investigación.

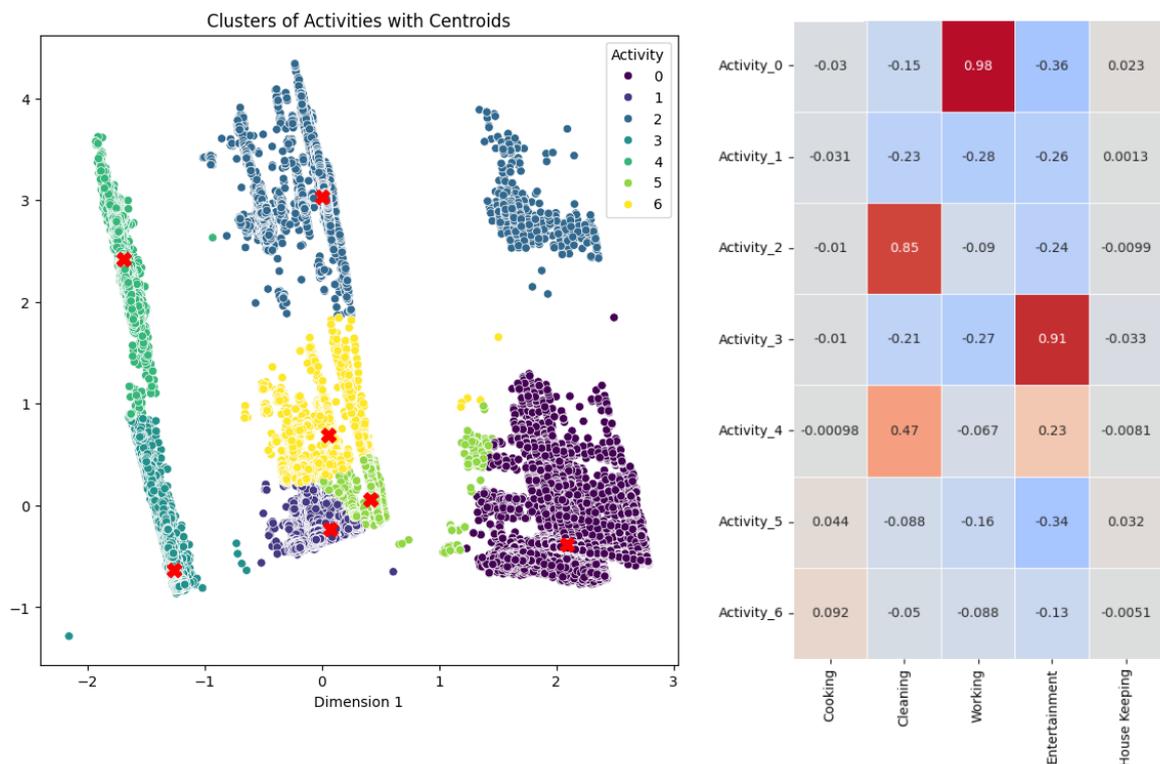


**Figura 5.6:** Serie temporal con el gasto de varios electrodomésticos a lo largo del día

### 5.4.3. Aplicación de técnicas de aprendizaje no supervisado

Tras la selección y preparación del *dataset* público REFIT, y la implementación de técnicas de limpieza, normalización y reducción de datos, se realizó un paso crucial en la preparación de los datos para su análisis posterior. Se agruparon los datos de los electrodomésticos según las actividades de la vida diaria que facilitan, como por ejemplo, combinar los datos del uso de microondas y tostadora bajo la actividad de "cocinar", integrada dentro de la escala fundamental de IADLs, como se estipula en el Capítulo 3. Este enfoque permite una detección más precisa y contextual de las actividades, al considerar la interacción con múltiples dispositivos como parte de una única actividad.

Con los datos ya agrupados por actividades del mismo tipo, el siguiente paso fue aplicar técnicas de aprendizaje automático no supervisado para clasificar las instancias basadas en sus similitudes. Los conocimientos necesarios para realizar este paso fueron en gran parte adquiridos en la asignatura de Desarrollo de Sistemas Inteligentes, impartida en el Máster de Ingeniería Informática. Entre los diversos posibles algoritmos a utilizar, se seleccionó de manera preliminar el algoritmo K-Means por su eficacia en la identificación de agrupaciones naturales dentro de los datos, y su eficiencia al tratar con gran cantidad de información, presumiendo un mejor rendimiento que otros algoritmos como DBSCAN o clustering jerárquico.



**Figura 5.7:** Aplicación y correlación del algoritmo K-Means para con los datos de electrodomésticos

Para optimizar el rendimiento de K-Means y adaptarlo a nuestro conjunto de datos, tal y como se aprecia en el Listado 5.1, se realizó una búsqueda exhaustiva de los hiperparámetros óptimos. Este proceso de afinación de hiperparámetros es esencial para mejorar la capacidad del algoritmo de formar agrupaciones significativas que reflejen con precisión las diferentes actividades de la vida diaria detectadas a través del uso de electrodomésticos y otros aparatos electrónicos domésticos. La aplicación de K-Means, junto con la selección de hiperparámetros, nos permitió agrupar las instancias en clústeres con gran afinidad a las actividades, como se puede comprobar con la matriz de correlación Fig. 5.7 que se realizó después del algoritmo. Las celdas de color rojo indican una gran correlación (en un rango entre -1 y 1, estará rozando el límite superior) entre fila y columna y, por tanto, se puede observar cómo encuentra tres clústeres, que en la imagen son las filas, yendo desde 0 hasta el número

de agrupaciones que encuentre menos uno, con gran correlación con las columnas correspondientes a 'limpieza, trabajo y entretenimiento'.

**Listado 5.1:** Hiperparametrización del algoritmo K-Means

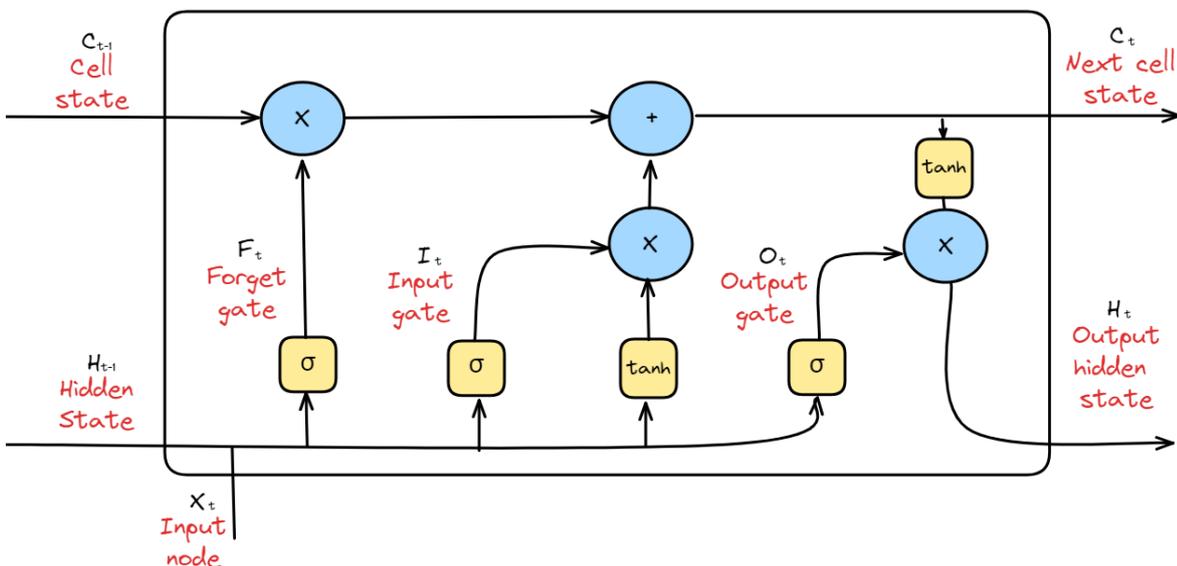
```

1 param_dist = {
2     'n_clusters': [3, 5, 7],
3     'init': ['k-means++', 'random'],
4     'max_iter': [300, 500, 700, 900],
5     'n_init': [5, 10, 15, 20],
6 }
7 kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
8 random_search = GridSearchCV(kmeans, param_dist, cv=3)
9 random_search.fit(reduced_data)

```

#### 5.4.4. Aplicación de técnicas de aprendizaje supervisado

Después de emplear el algoritmo K-Means para clasificar las instancias en distintos clústeres, correspondientes a varias actividades de la vida diaria, se generó un conjunto de datos etiquetado que nos sirve como 'ground truth' o verdad fundamental. Es decir, si en un clúster se puede establecer una relación entre una instancia y una actividad, o lo contrario, que una instancia no pertenece a ninguna actividad, esta información nos servirá como verdad fundamental. En la literatura, podemos encontrar varios artículos que utilizan aprendizaje no supervisado para clasificar información antes de usar aprendizaje supervisado [73, 74, 75]. Este conjunto de datos se utiliza ahora para entrenar un modelo de red neuronal más avanzado, específicamente una Red Neuronal de Memoria a Corto y Largo Plazo (LSTM por sus siglas en inglés), que está diseñada para manejar secuencias de datos temporales, lo cual es ideal para el contexto de monitorización de actividades de la vida diaria a través de dispositivos conectados en el hogar.



**Figura 5.8:** Estructura de una célula en una red LSTM

Las redes LSTM son particularmente adecuadas para este tipo de datos, ya que tienen la capacidad de aprender y recordar patrones a lo largo del tiempo [49], lo cual es esencial para identificar actividades basadas en la secuencia y duración de uso de diferentes electrodomésticos que envían información a la vez, y por eso ha sido elegida por encima de otro tipo de algoritmos como árboles de decisión (SVM) o modelos ocultos de Markov (HMM). Este enfoque nos permite no solo clasificar las instancias en actividades existentes, sino también predecir la ocurrencia de nuevas instancias de actividad basadas en patrones temporales de uso de energía, dividiendo el *dataset* en entrenamiento,

validación y test. Como podemos ver en Fig. 5.8, cada célula de las capas LSTM tiene un estado anterior, que puede o no olvidar dependiendo de la información que le entre, y propagando la información resultante a la siguiente célula.

El modelo LSTM específico que hemos diseñado incluye dos capas LSTM, y es también conocida como *Stacked* o *deep* LSTM, introducida por [76]. Cada una de estas capas LSTM está seguida de una capa de "Dropout" para reducir el sobre ajuste al descartar aleatoriamente algunas unidades de la red durante el entrenamiento, lo que mejora la generalización del modelo. La primera capa LSTM tiene 100 unidades y devuelve secuencias completas para proporcionar estos datos temporales a la siguiente capa. La segunda capa LSTM mantiene la estructura secuencial antes de la clasificación final. La capa de salida es una capa densa conectada a la última capa LSTM, con una activación 'softmax' para obtener una distribución de probabilidad sobre las posibles actividades de la vida diaria, facilitando así la clasificación multiclase.

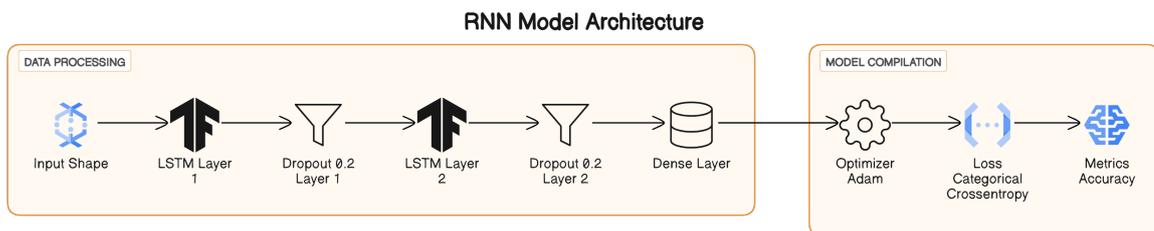


Figura 5.9: Capas de la red LSTM del proyecto

El modelo se compila con el optimizador 'adam', conocido por su eficiencia en el entrenamiento de redes neuronales, y se utiliza la pérdida de 'categorical\_crossentropy' adecuada para problemas de clasificación multiclase. La métrica de 'accuracy' nos permitirá evaluar el rendimiento del modelo en términos de su capacidad para clasificar correctamente nuevas instancias en la actividad correspondiente. Este es el código en Python para la definición del modelo:

Listado 5.2: Red neuronal LSTM

```

1 model = Sequential()
2 model.add(LSTM(units=100, return_sequences=True, ←
   ← input_shape=(input_shape, 1)))
3 model.add(Dropout(0.2))
4 model.add(LSTM(units=100))
5 model.add(Dropout(0.2))
6 model.add(Dense(units=output_shape, activation='softmax'))
7 model.compile(optimizer='adam', ←
   ← loss='categorical_crossentropy', metrics=['accuracy'])

```

Con este modelo LSTM, avanzamos hacia un sistema capaz de clasificar y predecir automáticamente las actividades de la vida diaria en hogares inteligentes, basándose en patrones de consumo de energía.

El enfoque descrito en esta iteración es útil bajo la condición de detección de actividades con más de un electrodoméstico/aparato electrónico enviando información en la misma instancia. En el caso de que la información llegara separada desde la base de datos, ya fuera por tipos de conjuntos de enchufes de la misma actividad, o por enchufes en solitario, habría que decidir si juntar la información y aplicar el modelo que se acaba de exponer en este trabajo, o intentar detectar actividades sin tenerse en cuenta entre ellas, lo cual podría generar incertidumbre, ya que se debe tener en cuenta que al tratarse de IADLs, y no de ADLs, el usuario puede no estar haciendo una sola actividad a la vez (por ejemplo, tener la televisión encendida mientras va a prepararse una comida al microondas), y es por ello que su manejo en conjunto ayuda a tener una visión más general de qué está haciendo la persona mayor en cada momento.

### 5.4.5. Modelo único Vs. Modelo individual

En los meses previos a agosto, se desarrolló un modelo de red neuronal utilizando datos agregados de varias casas para detectar actividades de la vida diaria. Este enfoque global ofreció una comprensión general de los patrones de consumo de energía y su relación con las actividades diarias. Sin embargo, tras un análisis de los resultados y considerando la variabilidad inherente en los comportamientos y rutinas individuales de cada hogar, se hizo evidente la necesidad de adoptar un enfoque más personalizado.

Reconociendo que cada hogar tiene sus propias características únicas y que las personas interactúan con sus dispositivos de manera diferente, se decidió avanzar hacia la creación de modelos individualizados para cada hogar. Este enfoque permite una mayor precisión en la detección de actividades, ajustando el modelo a las especificidades y rutinas únicas de cada familia, lo que resulta en una mejor adaptabilidad y eficacia del sistema.

Para implementar esta estrategia, desarrollamos un código flexible capaz de generar y entrenar un modelo específico para cada hogar. Este código automatiza el proceso de segmentación de datos, entrenamiento y validación de modelos, asegurando que cada modelo esté finamente ajustado a los patrones de consumo de energía del hogar correspondiente. La estructura del código se diseñó para ser escalable y fácilmente adaptable, permitiendo la incorporación de nuevos datos y la actualización de modelos a medida que se recopila más información.

Este enfoque individualizado mejora la precisión de la detección de actividades en cada hogar. Al adaptar los modelos a las características específicas de cada hogar, podemos proporcionar información más precisa y acciones que se alinean estrechamente con las necesidades y comportamientos reales de los individuos, mejorando así la experiencia del usuario y la efectividad general del sistema.

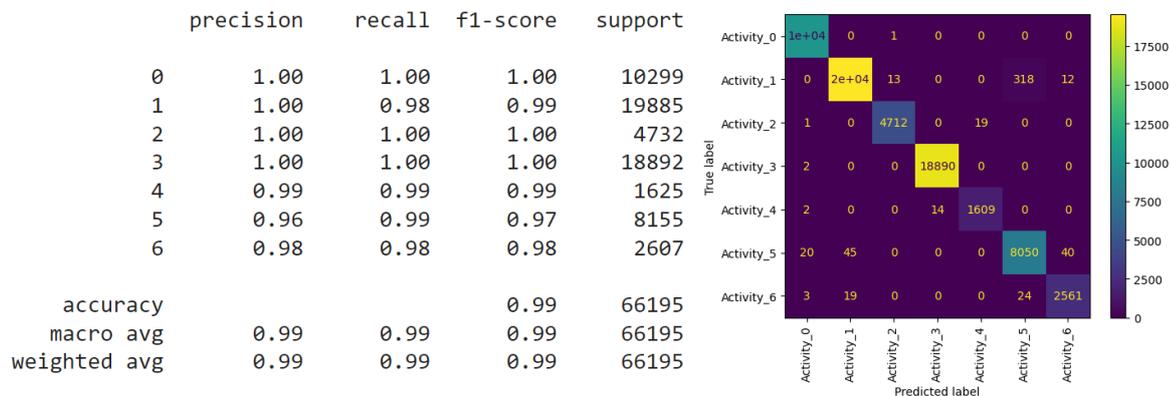


Figura 5.10: Precisión de la red neuronal LSTM

En la Figura 5.10 se puede ver cómo en el eje Y están los valores verdaderos de cada instancia, agrupados por cada actividad detectada por el algoritmo K-Means, y en el eje X están los valores que ha predicho la red neural para cada instancia, también agrupados por actividades (Actividad 0 a 6). Los valores que no estén en la diagonal de la matriz de correlación son valores que se han predicho de manera errónea, siendo estos la minoría en comparación con los que se han acertado.

### 5.4.6. Evaluación del modelo

La evaluación de nuestro modelo, basado en técnicas avanzadas de aprendizaje automático, ha arrojado resultados prometedores que subrayan la viabilidad de detectar actividades de la vida diaria a través del monitoreo inteligente en hogares. Al aplicar métodos hiperparametrizados a nuestros conjuntos de datos, hemos podido observar cómo el modelo aprende de las instancias presentes en múltiples hogares, y puede predecir las actividades en nuevas instancias, demostrando capacidad notable para adaptarse a las peculiaridades del entorno doméstico. Esta adaptabilidad es crucial, ya que reconoce la diversidad en los patrones de comportamiento humano y permite, por tanto, definir

aquellos comportamientos que se podrían definir como anómalos. Los resultados obtenidos muestran el poder del aprendizaje automático para transformar la comprensión y enfoque hacia la creación de entornos inteligentes.

#### 5.4.7. Resultados de la iteración

Los primeros resultados de este trabajo fueron presentados [77] con éxito en noviembre de 2023 en Riviera Maya, México, durante la conferencia internacional *UCAmI (Ubiquitous Computing and Ambient Intelligence<sup>2</sup>)*, un congreso centrado en avances en computación ubicua e inteligencia ambiental. En la presentación se expusieron los hallazgos preliminares, hasta esta tercera iteración, permitiendo el intercambio de ideas y retroalimentación con expertos en el campo.

A partir de este trabajo, se propone el uso de un mayor número de algoritmos de aprendizaje no supervisado y comparar sus resultados con K-Means, para obtener así una comparación empírica del porqué esta técnica era la más adecuada en este proyecto.

#### 5.4.8. Mejoras

Se procedió a implementar una comparación con otros posibles algoritmos populares. *DBSCAN* [78], pese a ser una opción, se descartó por su interminable tiempo de compilación con los recursos disponibles, el cual hacía totalmente inviable cualquier intento de usarlo en una cantidad de datos grande como esta. El *Clustering Jerárquico* [79], pese a que también presenta una gran complejidad y tiempo de ejecución, y un Modelo de Mezcla Gausiano (GMM) [30], fueron las opciones finales seleccionadas. En la Tabla 5.2 podemos ver los resultados comparativos a través de varias técnicas de validación de estos algoritmos ejecutados sobre el mismo conjunto de datos. Aparte, en la Figura 5.11 podemos ver cómo agrupan las mismas instancias, de manera distinta, en un número de conjuntos mayor o menor.

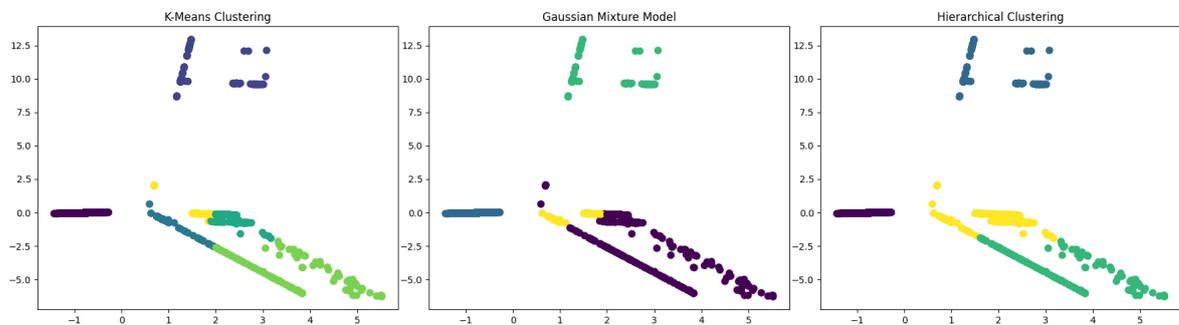


Figura 5.11: Comparación los algoritmos de K-Means, GMM y clustering jerárquico

Tabla 5.2: Comparación de los tres algoritmos no supervisados

Algoritmo	Silhouette	Davies-Bouldin	Calinski-Harabasz	Tiempo
K-Means	0,8796	0,2873	864.238,0685	913,942 s
GMM	0,9146	0,1263	507.246,6264	163,618 s
Jerárquico	0,9188	0,3335	595.612,2949	1.214,018 s

Las conclusiones a extraer vienen precedidas por varios factores a tener en cuenta. La hiperparametrización del clustering jerárquico y del K-Means hacen que su tiempo de ejecución sea mucho mayor que el del GMM. Pese a que en las técnicas de validación los resultados son muy parejos entre los tres, hay que irse a la matriz de correlación de la Figura 5.12 para darse cuenta del porqué es importante la hiperparametrización, ya que en el GMM no se detecta 1 de las 4 actividades (la de limpieza), al contrario que en las otras 2.

<sup>2</sup><https://ucami.org/>

En conclusión, es complicado decidir cuál es el mejor algoritmo, y dependerá de los datos que se analicen, pero en general los resultados serán parejos entre el K-Means y el Clustering Jerárquico, pese a que tarden bastante en ejecutarse. Por tanto, dentro del código se ejecutarán todas las opciones y aquella que de mejores resultados en las técnicas de validación será la elegida.

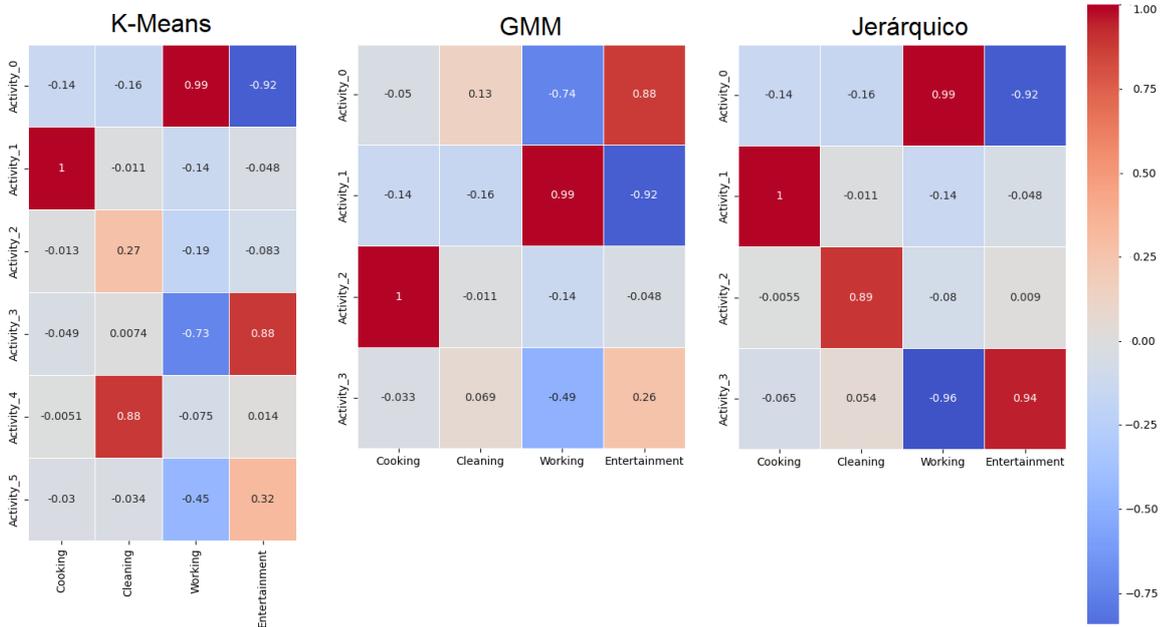


Figura 5.12: Matriz de correlación de los algoritmos

## 5.5. ITERACIÓN 4: COMPONENTE SOFTWARE RUST PARA LA RECOGIDA DE INFORMACIÓN

### 5.5.1. Creación del componente software

Para integrar y gestionar eficazmente la interacción con los enchufes inteligentes TAPO P110, se desarrolló un componente utilizando una librería en Rust, un lenguaje de programación conocido por su seguridad y eficiencia. Esta API se diseñó con el objetivo de solicitar y recopilar únicamente la información relevante de los dispositivos que están registrados y conectados a la misma red que el servidor o intermediario que ejecuta el código.

El enfoque de este código se centra en asegurar que la comunicación con los enchufes inteligentes sea directa y se limite a los datos necesarios para el análisis de actividades y anomalías dentro de los hogares inteligentes, como se puede ver en el diagrama de flujo del software, en la Fig. 5.13. Al emplear Rust, se garantiza que la interacción con los dispositivos no solo sea rápida y confiable, sino también segura, minimizando los riesgos asociados con la gestión de dispositivos IoT en redes domésticas.

El diagrama muestra el funcionamiento del sistema de monitoreo de energía y comienza con la configuración inicial donde se carga la clave que nos permite la conexión con la base de datos MongoDB desde una variable de entorno. Se configura un sistema de registro (*logging*) para seguir el flujo de las actividades del sistema e identificar más fácilmente los errores.

A continuación, el sistema procede a leer las variables de entorno necesarias para la autenticación con los dispositivos Tapo. Utiliza esta información para descubrir las direcciones IP de los dispositivos Tapo en la red y crea clientes para cada uno de ellos, permitiendo la comunicación y el control de los dispositivos, como se puede apreciar en el Listado 5.3.

Una vez establecida la comunicación, el sistema inicia un bucle de monitoreo. En cada iteración del bucle, se recopilan datos de cada dispositivo en la lista. Para cada dispositivo, el sistema verifica

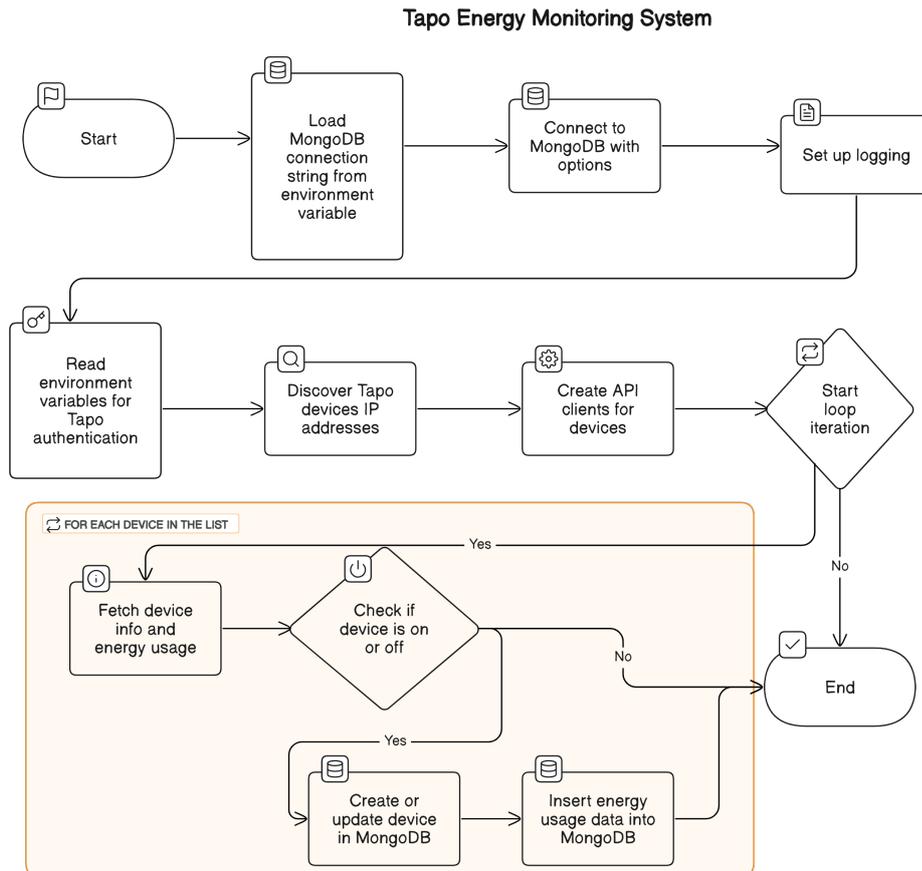


Figura 5.13: Diagrama de flujo de la API

si está encendido o apagado. Si el dispositivo está encendido, el sistema recoge información del dispositivo y datos sobre su consumo de energía. Esta información se utiliza para crear o actualizar registros en la base de datos MongoDB, donde después se insertan los datos de consumo de energía.

Listado 5.3: Código para la conexión con los enchufes

```

1 println!("Starting IP discovery...");
2 let discovered_ips = discover_tapo_devices();
3 println!("Discovered IPs: {?}", discovered_ips);
4
5 // Discover devices
6 let device_futures: Vec<_> = discovered_ips.iter()
7 .map(|ip| ApiClient::new(tapo_username.clone(), ↵
8   ↵ tapo_password.clone()).p110(ip.clone()))
9 .collect();
10
11 let devices: Vec<Result<PlugEnergyMonitoringHandler, ↵
12   ↵ tapo::Error>> = ↵
13   ↵ futures::future::join_all(device_futures).await;
14 println!("API Clients created for {} devices.", devices.len());
  
```

Si el dispositivo está apagado, el sistema simplemente continúa con el siguiente dispositivo en la lista sin realizar ninguna acción. Este proceso se repite para cada dispositivo, permitiendo una supervisión continua y en tiempo real del consumo de energía.

El proceso finaliza una vez que todos los dispositivos han sido explorados y los datos necesarios han sido almacenados en la base de datos, cerrando el ciclo de monitoreo hasta la próxima iteración, la cual ocurre pasados unos segundos. Este sistema asegura que se mantenga un registro detallado del consumo de energía, facilitando el análisis basado en los datos recogidos.

### 5.5.2. Elección base de datos

Para el almacenamiento de la información relevante extraída de los enchufes TAPO P110, como se comentó en la Iteración 2, se utilizó una base de datos NoSQL adecuada que cumpliera con los requisitos específicos del proyecto. Inicialmente, se realizaron pruebas con Firebase debido a su facilidad de uso, escalabilidad en tiempo real y eficiente integración con aplicaciones web y móviles. Sin embargo, tras una evaluación, se optó por MongoDB como la solución definitiva.

La decisión de elegir MongoDB sobre Firebase se basó en varios factores clave:

1. **Esquema Flexible:** MongoDB ofrece una estructura de datos más flexible que permite almacenar documentos sin un esquema fijo. Esto resulta particularmente ventajoso para nuestro proyecto, dado que la variedad y naturaleza de los datos recogidos de los enchufes inteligentes pueden cambiar o evolucionar con el tiempo.
2. **Escalabilidad y Rendimiento:** Aunque Firebase proporciona una buena escalabilidad, MongoDB destaca en la gestión de grandes volúmenes de datos y cargas de trabajo intensivas, ofreciendo un rendimiento superior en operaciones de lectura y escritura, lo cual es crucial para el análisis en tiempo real de los datos de los enchufes.
3. **Consultas más Poderosas:** MongoDB admite consultas más complejas y variadas, lo que permite un análisis más profundo y detallado de los datos. Esta capacidad es esencial para extraer resultados significativos como las actividades y las anomalías.
4. **Control y Propiedad de los Datos:** Al utilizar MongoDB, se tiene un mayor control sobre la infraestructura de datos y su gestión. Esto contrasta con Firebase, donde los datos residen en la infraestructura de Google, lo que podría plantear preocupaciones de privacidad y dependencia de un proveedor.
5. **Costo y Escalabilidad a Largo Plazo:** MongoDB ofrece opciones más flexibles y baratas para la escalabilidad a largo plazo, especialmente cuando el proyecto requiere un almacenamiento de datos considerable y operaciones de alta complejidad.

Por estas razones, MongoDB, aparte de ser una base de datos con la cual se tenía experiencia previa, se presentó como la opción más adecuada para nuestro proyecto, proporcionando la robustez, flexibilidad y escalabilidad necesarias para gestionar eficazmente los datos recogidos de los enchufes TAPO P110 y facilitar análisis complejos y personalizados de las actividades domésticas.

Dentro de la infraestructura de almacenamiento de datos de nuestro proyecto, se ha estructurado la base de datos MongoDB utilizando dos colecciones para organizar y gestionar la información recopilada de los enchufes inteligentes. Esta organización facilita el acceso y el análisis de los datos para la detección de comportamientos anormales, permitiendo una gestión detallada y el procesamiento de la información, como se puede apreciar en la Fig. 5.14.

- **Colección 'devices':** Esta colección se dedica a almacenar información básica sobre los enchufes inteligentes en sí. Cada documento dentro de la colección 'devices' contiene el identificador único del enchufe, que sirve como clave principal para referenciar el dispositivo dentro de la base de datos. Además, para una mejor identificación posterior a la hora de mostrar la información en posibles *front-ends*, y dado que las iteraciones y modelos anteriores estaban más centradas en la detección de IADLs, pero a partir de esta iteración el trabajo se centró más en la detección de anomalías, el nombre del enchufe al instalarlo en la aplicación deberá tener el siguiente formato predefinido: **ModeloEnchufe-Usuario-Habitación-Electrodoméstico**. Por ejemplo, 'P110-Usuario1-Cocina-Microondas'. De esta manera al estar más centrada esta iteración en la futura detección de comportamientos anómalos que en la detección de actividades, la cual en este caso se hará de manera implícita entre el enchufe y la actividad, permitirá una detección de anomalías más sencilla al poder saltar el paso anterior y toda la incertidumbre que le rodeaba, y ayudará en una búsqueda más precisa de estos *outliers*.

- Colección 'data':** La colección 'data' está diseñada para capturar y almacenar las instancias de datos recopiladas de los enchufes inteligentes. Cada documento en esta colección incluye varios campos para el análisis de la información: la fecha y hora de cada instancia registrada; una referencia al dispositivo correspondiente, estableciendo una relación con la colección 'devices'; el estado del dispositivo (es decir, si el electrodoméstico está encendido o apagado en el momento de la instancia); la potencia que está consumiendo el electrodoméstico, lo cual es fundamental para el análisis del consumo energético; y el identificador único de la instancia, asegurando la unicidad y la trazabilidad de cada uno de los datos. Aparte, por motivos que se explicarán en la siguiente iteración, de manera posterior a esta se añadió un nuevo campo: 'synthetic', el cual indicará si el *crate* ha sido adquirido manualmente o creado artificialmente.

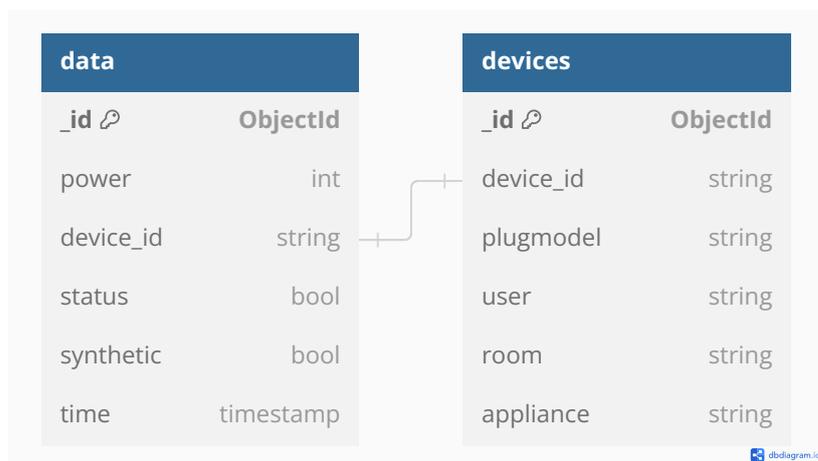


Figura 5.14: Esquema de la base de datos

### 5.5.3. Elección de gateways

Para la monitorización eficiente de las anomalías en hogares mediante el uso de enchufes inteligentes TAPO P110, se implementó una solución que utiliza una Raspberry Pi como puente entre los enchufes y la red. Este enfoque no solo aprovecha la portabilidad de las Raspberry Pi, sino que también minimiza lo intrusivo del proyecto, al ser dispositivos de pequeño tamaño que se integran discretamente en el entorno doméstico.

La Raspberry Pi, un pequeño ordenador y de bajo costo, se ha seleccionado por su capacidad para ejecutar una amplia gama de aplicaciones, su bajo consumo energético y su tamaño reducido, lo que la hace ideal para este tipo de aplicaciones domésticas. Cada Raspberry Pi se configura con la API desarrollada en Rust. El proceso se estructura de la siguiente manera:

- Configuración de Raspberry Pi:** Cada Raspberry Pi se configura con el software necesario, sistema operativo, y las librerías necesarias, y se conecta a la red doméstica. Esto permite que la Raspberry Pi actúe como un intermediario entre los enchufes inteligentes y la red.
- Instalación de la API Rust:** La API Rust se instala en cada Raspberry Pi, proporcionando una interfaz para comunicarse directamente con los enchufes inteligentes TAPO P110. La API está diseñada para solicitar y recopilar solo la información relevante de los dispositivos, optimizando así el flujo de datos.
- Minimizando la Intrusividad:** Al utilizar dispositivos de pequeño tamaño como las Raspberry Pi, el proyecto minimiza su impacto visual y físico en el hogar. Se debe encontrar un lugar en el hogar que permita una conexión rápida, pero que no resulte molesto para el usuario y, por tanto, sea poco intrusivo visualmente.
- Recolección de Datos:** Una vez se declaran las variables de entorno que permiten autenticarse con los enchufes y con la base de datos, se ejecuta el código para que comience la recolección

ininterrumpida de datos de los enchufes inteligentes, los cuales lo transmiten a través de internet para su análisis posterior. Esta configuración permite un monitoreo continuo de las actividades y anomalías en el hogar, sin necesidad de una intervención manual.

5. **Análisis y Monitoreo:** Los datos recogidos se analizan para identificar anomalías en las actividades de la vida diaria, permitiendo un monitoreo eficaz y proporcionando resultados relevantes sobre el comportamiento de los residentes.

Este enfoque no solo garantiza una recopilación de datos eficiente y segura, sino que también respeta la privacidad y el confort de los habitantes del hogar, al utilizar una tecnología que se integra suavemente en su entorno cotidiano. La combinación de Raspberry Pi y la API Rust representa una solución efectiva para el monitoreo inteligente de hogares, alineándose con el objetivo principal del proyecto.

#### 5.5.4. Empaquetado del programa

En un esfuerzo por mejorar la accesibilidad y facilitar la implementación en dispositivos con recursos computacionales limitados (como Raspberrys antiguas), se ha optado por dockerizar este código (implementar Docker, una herramienta para la automatización del despliegue de aplicaciones, para poder empaquetar el software en un contenedor). Se pretende así encapsular el programa y todas sus dependencias en un contenedor independiente del sistema operativo del dispositivo donde se ejecuta. Esto permite que el programa funcione de manera idéntica en cualquier plataforma que soporte Docker, eliminando así los problemas de compatibilidad y las complejidades asociadas a la instalación y configuración del software.

Para facilitar la actualización del contenedor en Docker Hub,<sup>3</sup> un repositorio de contenedores, se ha utilizado una plataforma de automatización de flujo de trabajo como GitHub Actions<sup>4</sup>. Este sistema está diseñado para actualizar automáticamente la imagen en Docker Hub cada vez que se realiza un *commit* al repositorio donde reside el código del programa, cuyo enlace se encuentra en el Anexo A. El flujo de trabajo (*workflow*) de GitHub Actions está configurado para construir la imagen del Docker, etiquetarla adecuadamente y luego empujarla a Docker Hub, asegurando que cualquier mejora o corrección de errores se propague rápidamente a todos los usuarios finales.

En el *Dockerfile*, un archivo de texto que contiene una serie de instrucciones necesarias para crear una imagen con Docker, se instalarán las dependencias Rust necesarias dentro del contenedor creado, y se ejecutará el código una vez pasadas las variables necesarias (nombres de usuario y contraseñas) a través de la línea de comandos.

#### 5.5.5. Despliegue preliminar

Para llevar a cabo las pruebas iniciales del código diseñado para interactuar con los enchufes propio hogar, donde reside una persona mayor sola. Este entorno se consideró ideal para evaluar la efectividad y cómo de práctica es la solución en un contexto real de monitoreo de actividades y anomalías, especialmente en un ambiente habitado por personas mayores que podrían beneficiarse significativamente de la implementación de tecnologías asistidas.

Es importante destacar que, a pesar de la naturaleza sensible del monitoreo en hogares con personas mayores, la metodología empleada para las pruebas de este programa han sido diseñadas para asegurar la máxima protección de la privacidad. La configuración del código permite recoger únicamente datos relacionados con el consumo energético y la operación de los dispositivos conectados a los enchufes inteligentes, sin recabar ningún tipo de información personal o datos que puedan identificar directa o indirectamente a la persona residente, ni almacena ningún dato relacionado con la salud o el estado físico de los individuos en el hogar. La información recogida se limita estrictamente a parámetros operacionales y de consumo energético de los dispositivos conectados. En la Figura

---

<sup>3</sup><https://hub.docker.com/>

<sup>4</sup><https://docs.github.com/en/actions>



(a) Monitor en un entorno de laboratorio

(b) Microondas en un entorno de laboratorio

**Figura 5.15:** Raspberry recogiendo datos de los enchufes inteligentes

5.15 se comprueba que el funcionamiento de la recolección de datos y la difusión y almacenamiento funcionan correctamente.

Estas medidas garantizan que, a pesar de la naturaleza detallada del monitoreo realizado, la privacidad de la persona mayor en el hogar se mantiene intacta, y no se produce ninguna brecha en la privacidad. La implementación cuidadosa y considerada del programa, junto con las estrictas salvaguardas de privacidad, aseguran que las pruebas se realicen de manera ética y respetuosa, alineándose con los principios de responsabilidad y protección de datos personales. La evaluación con un despliegue real en distintos hogares se llevará a cabo en un periodo no muy lejano de tiempo, una vez el comité de ética apruebe la solicitud de evaluación en entornos reales para este trabajo de investigación.

## 5.6. ITERACIÓN 5: MODELO PARA LA DETECCIÓN DE ANOMALÍAS EN LA IADLS

Los datos temporales, especialmente aquellos que siguen un ciclo como las horas del día, presentan un desafío particular: el final y el principio del ciclo son puntos contiguos, no separados por una gran diferencia numérica. Es decir, aunque numéricamente las 23:00 parecen lejanas a las 00:00, en realidad están muy próximas, ya que marcan la transición de un día al siguiente. Esta peculiaridad puede complicar análisis que dependen de distancias numéricas lineales, como algunos modelos estadísticos o algoritmos de clustering, que podrían interpretar incorrectamente la proximidad de eventos que ocurren al final y al inicio del día. Por ello, es necesario una representación circular de estos datos que ayude a analizar esta continuidad natural en los datos a lo largo del tiempo.

Como una posible solución, en este proyecto las horas del día se transforman en coordenadas sinusoidales y cosinusoidales utilizando las siguientes fórmulas:

$$\text{hour\_sin} = \sin\left(\frac{2\pi \times \text{HoraDelDía}}{24}\right) \quad (5.1)$$

$$\text{hour\_cos} = \cos\left(\frac{2\pi \times \text{HoraDelDía}}{24}\right) \quad (5.2)$$

Esta transformación mapea cada hora a un punto en el círculo unitario, pasando el problema de un espacio unidimensional (hora del día) a un espacio bidimensional circular. Esto permite que distancias y relaciones sean evaluadas de manera que respeten la continuidad del ciclo diario, facilitando la identificación de patrones o anomalías que respetan la naturaleza cíclica de los datos.

A partir de aquí, una vez representados los datos de manera continua y cíclica, hará falta encontrar aquellos valores anómalos. Revisando las técnicas que se suelen utilizar en estos problemas, una posible solución es el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise), ya mencionado en anteriores iteraciones. Este algoritmo es particularmente adecuado para

datos donde las anomalías se manifiestan como puntos aislados o grupos pequeños en un espacio que de otro modo sería densamente poblado. Las características clave de DBSCAN que lo hacen adecuado son:

1. **Capacidad para identificar 'ruido':** DBSCAN clasifica puntos como *core points*, *border points* o *noise points*. Los puntos etiquetados como 'ruido' pueden interpretarse directamente como anomalías.
2. **No requiere un número predeterminado de clusters:** A diferencia de K-means, DBSCAN no necesita que el usuario especifique el número de clusters de antemano, lo que es ideal para datos donde el número de clusters (o anomalías) no es conocido de antemano.
3. **Sensibilidad a la densidad local:** Con los parámetros 'eps' (la máxima distancia entre dos muestras para que se consideren en el mismo vecindario) y 'min\_samples' (el número mínimo de muestras en un vecindario para que un punto sea considerado como un punto central), DBSCAN puede adaptarse a variaciones en la densidad de los datos, lo que es crucial en datos temporales donde la actividad puede variar significativamente a lo largo del día. Hacer un ajuste correcto de estos dos parámetros será clave para poder identificar las anomalías significativas.

Este enfoque permite que DBSCAN maneje bien los datos con clusters de formas irregulares y tamaños variables, así como la capacidad de identificar puntos de datos que no pertenecen a ningún cluster, clasificándolos efectivamente como anomalías.

La combinación de una representación circular con un algoritmo de clustering basado en densidad como DBSCAN (ejemplo en la Figura 5.18 al final del Capítulo) proporciona una metodología simple, pero efectiva para identificar comportamientos anómalos en series temporales que siguen patrones cíclicos, haciendo esta técnica útil para análisis en el dominio del monitoreo de personas mayores.

### 5.6.1. Validación del sistema de detección de IADL y comportamientos anormales mediante gemelos digitales

Con el objetivo de generar una mayor validación de los modelos creados, se usó la aplicación SiSHoDit [80] creada con el motor de videojuegos Godot para recrear un gemelo digital de un hogar que permite a un agente el generar ADLs, como se puede ver en la Figura 5.16.

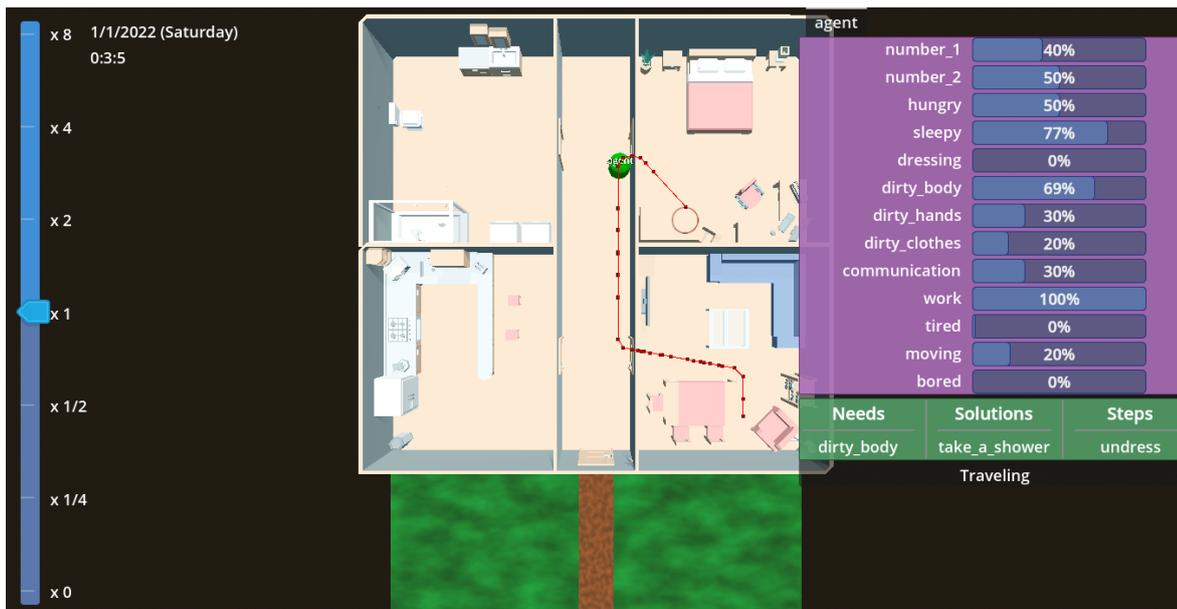


Figura 5.16: Herramienta SiSHoDit en funcionamiento

En este hogar virtual existían electrodomésticos con los que el agente podía y puede interactuar, tales como microondas, lavadoras, secadoras, televisores u ordenadores. Por tanto, generar registros

sintéticos a analizar por los modelos creados en este proyecto requería solo de pocos cambios en el sistema, y es por ello que se realizaron para generar esta información sintética con el objetivo de realizar esa validación y aumentar la cantidad de datos con los que se pruebe el modelo.

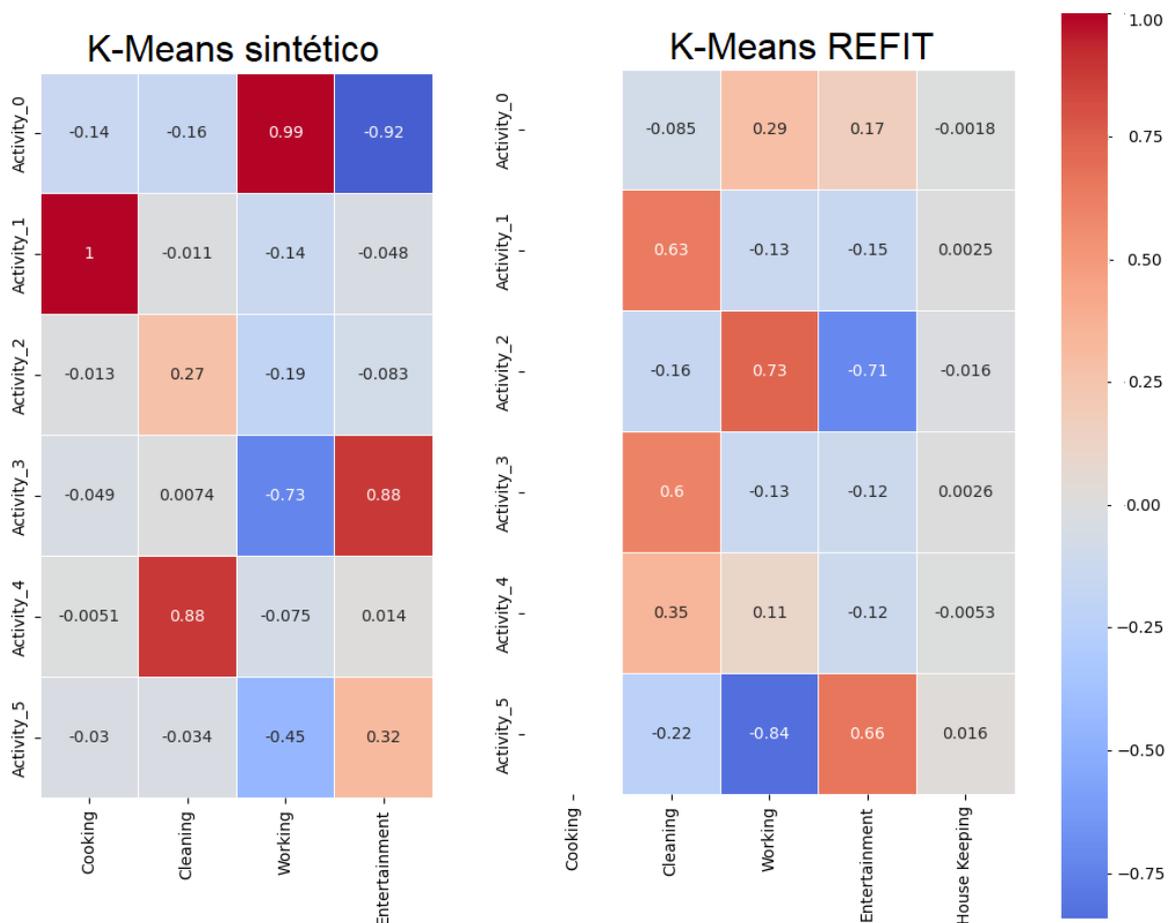
El formato de salida de los datos generados en un *log* JSON por la aplicación debían seguir el siguiente formato:

**Listado 5.4:** Formato de logging de la aplicación SiSHoDit

```
1 {"from": "washer-energy", "time": 150, "type": "energy", "value": 0},
2 {"from": "dryer-energy", "time": 15, "type": "energy", "value": 0},
```

Indicando el electrodoméstico que emite la instancia, el tiempo en segundos desde que inició la simulación (lo cual nos permitirá luego pasar a fecha y hora, ya que se sabe la hora de inicio de la simulación), el 'tipo' cuyo resultado siempre será 'energía' (hay otros tipos para sensores de presencia, por ejemplo, que no son relevantes para este trabajo), y por último, el valor en vatios del electrodoméstico. Para cada electrodoméstico, se realizó un estudio real de su funcionamiento utilizando los enchufes Tapo para poder recrear con la mayor fidelidad posible los resultados en el gemelo digital.

Estos datos luego serían almacenados en la misma base de datos MongoDB que los obtenidos a través del intermediario Rust creado en la anterior iteración, con el distintivo '*synthetic*', indicando que son generados artificialmente, y no recogidos en hogares reales. Para su uso en los modelos, se transforman al formato CSV y después se les pasan a los modelos.



**Figura 5.17:** Comparación matriz hogar sintético con hogar del REFIT

El gemelo digital permite una prueba controlada y repetible sin los costes o inconvenientes que conlleva la recopilación de datos en entornos reales. Al simular las interacciones en un ambiente

virtual, se pueden ajustar rápidamente parámetros y situaciones para observar cómo responden los modelos bajo diferentes condiciones o duraciones, facilitando un análisis de la robustez y sensibilidad del modelo. Este enfoque también ofrece la ventaja de poder generar datos a gran escala, lo que es esencial para entrenar modelos de aprendizaje automático con la complejidad y diversidad necesarias para que sean efectivos en escenarios del mundo real.

Los resultados de comparar los datos sintéticos con los que provienen del conjunto de datos de prueba *REFIT* y presumiblemente con el conjunto procedente de pruebas reales a realizar, muestran que los primeros tienden a sobre-entrenar el modelo debido a la falta de ruido que provoca un simulador tan ajustado, a contrario que en la vida real, donde existirá más ruido a la hora de identificar actividades, como se puede ver en la Figura 5.17. En esta Figura, dentro de la imagen superior se muestra una representación circular de las horas del día, similar a un reloj, pero con marcas de hora desde las 00:00 hasta las 23:00. La línea gruesa azul, formada por todas las instancias de la actividad, indica la densidad o frecuencia de trabajo en diferentes horas del día. También se observan algunos puntos rojos ubicados alrededor, que representan outliers o comportamientos anómalos en la actividad. En la imagen inferior de este gráfico se puede ver un diagrama de dispersión que muestra las ocurrencias de instancias anómalas en relación con las fechas y las horas del día, a lo largo de un período de tiempo específico. Los puntos rojos en este gráfico señalan las ocurrencias de comportamientos anómalos en momentos específicos, con mayor concentración de estos alrededor de las primeras horas del día.

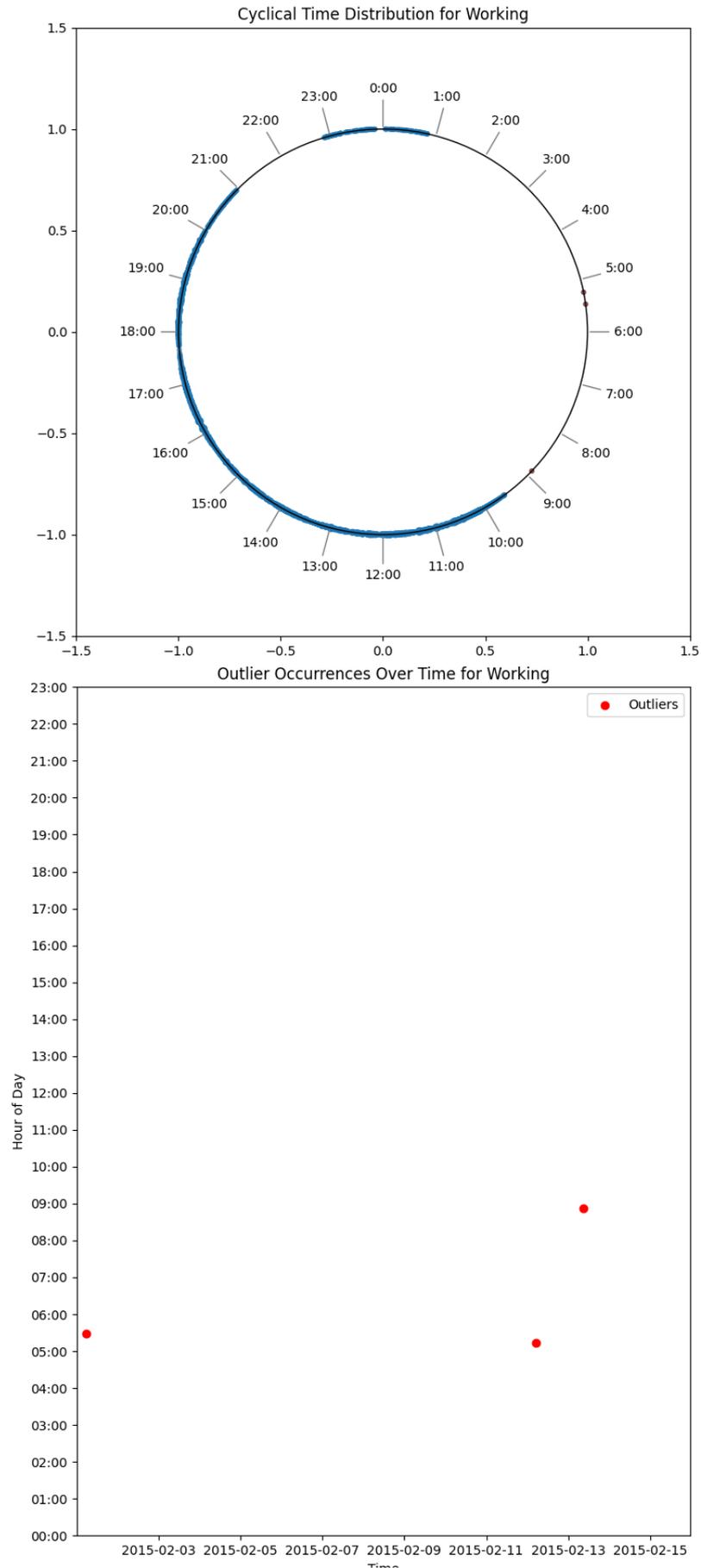


Figura 5.18: Valores anómalos en la actividad a lo largo de un mes



# Conclusiones

---

En este capítulo se realizará un juicio crítico y discusión sobre los resultados obtenidos. Aparte, se justificarán las competencias del máster adquiridas y se listarán posibles mejoras para un futuro inmediato en el proyecto.

## 6.1. DISCUSIÓN

La monitorización del consumo energético presenta un método innovador para la detección de actividades y anomalías en las Actividades Instrumentales de la Vida Diaria (IADLs). Sin embargo, es importante reconocer que este enfoque tiene limitaciones en cuanto a la gama de actividades que se pueden monitorear efectivamente. A diferencia de los métodos de monitorización más directos, como el uso de sensores de movimiento o cámaras, que pueden capturar una amplia variedad de actividades humanas, la monitorización basada en el consumo energético está fuertemente limitada a aquellas actividades que implican el uso de dispositivos eléctricos. Esto excluye una serie de IADLs, como la capacidad para manejar finanzas, la toma de medicamentos, o la interacción social, por ejemplo, que no necesariamente resultan en un consumo energético detectable.

Además, el método de monitorización basado en el consumo energético introduce desafíos significativos relacionados con la incertidumbre y los falsos positivos. Dado que este enfoque no requiere interacción humana directa para la detección de actividades, puede ser difícil discernir entre diferentes tipos de comportamientos, basándose únicamente en el perfil de consumo energético. Del mismo modo, y como ya se ha mencionado previamente, actividades simultáneas que consuman energía pueden confundirse, pudiendo llevar al algoritmo a interpretaciones erróneas sobre qué actividad se está realizando realmente. Esta superposición de actividades puede complicar la precisión de la detección de actividades, incrementando la tasa de falsos positivos.

La detección de anomalías en las IADLs a través del consumo energético introduce una capa adicional de incertidumbre, principalmente porque se basa en datos que ya de por sí pueden ser imprecisos. Además, los algoritmos utilizados para detectar estas anomalías no están exentos de errores y pueden generar falsos positivos. Sin embargo, en este proyecto, se ha intentado reducir esta incertidumbre vinculando directamente el uso de los enchufes a actividades específicas. Al hacer esto, se puede correlacionar más precisamente las instancias temporalmente inusuales en el consumo energético con actividades anómalas, mejorando la precisión en la detección de anomalías.

Por lo tanto, aunque la monitorización del consumo energético ofrece ventajas significativas en términos de discreción y novedad, será importante desarrollar estrategias complementarias que ayuden a mitigar estas limitaciones. Por ejemplo, la colaboración con los usuarios finales para establecer parámetros de normalidad y configurar personalmente los sistemas de alerta puede mejorar la precisión y la relevancia de la detección de actividades y anomalías.

## 6.2. OBJETIVOS ALCANZADOS

El objetivo general de este trabajo era *llegar a la detección de comportamientos anómalos en personas mayores que viven solas en el hogar utilizando dispositivos conectados, a través de su relación con las Actividades de la Vida Diaria, facilitando así la detección temprana de problemas que impacten negativamente en su autonomía y su calidad de vida*. Tras la realización de las 5 iteraciones, se ha conseguido inferir con precisión qué actividades se llevan a lo largo del día en el hogar del usuario, y detectar de manera simple anomalías en los datos, lo cual podrá permitir identificar comportamientos sospechosos en la persona mayor. Con respecto a los objetivos parciales:

- **Diseñar un marco de adquisición de datos para el establecimiento de patrones ad-hoc, relacionados con rutinas y comportamientos característicos:** Se ha desarrollado un marco que permite la adquisición de datos, definiendo un intermediario entre los enchufes inteligentes y la red, siendo este intermediario el que se encargará de todas las funciones que permitirán la conexión y recolección de datos. En las pruebas realizadas, se usan Raspberry Pi como intermediarias, debido a su capacidad para poder llevar estas operaciones sin presentar mayor inconveniente, demostrando que el marco funciona y se ajusta correctamente a las necesidades del proyecto.
- **Monitorizar las interacciones del usuario con los dispositivos conectados en el entorno (monitorización de bajo nivel), incluyendo la adquisición, comunicación y preprocesamiento de los datos, y su relación con el marco propuesto:** La monitorización de bajo nivel se ha logrado mediante la recolección de datos detallados de consumo de energía de los enchufes inteligentes, que reflejan directamente las interacciones del usuario con los dispositivos conectados. El uso de una API, y la infraestructura basada en puentes o *gateways* que permiten la adquisición, comunicación a una base de datos y el preprocesamiento de estos datos, asegura que solo la información relevante sea recopilada y transmitida para su análisis posterior, en línea con el marco propuesto en el anterior objetivo parcial.
- **Extraer conocimiento de los datos adquiridos (o a partir de un conjunto de datos públicos relacionados), mediante el empleo de algoritmos Machine Learning de aprendizaje supervisado y no supervisado, tanto para el análisis de las rutinas comportamentales como para la detección de situaciones anómalas, y su relación con las Actividades de la Vida Diaria:** El uso de algoritmos de Machine Learning, tanto de aprendizaje supervisado como no supervisado, para analizar los datos recogidos ha permitido extraer conocimientos significativos sobre las actividades que se desarrollan en los hogares diariamente. La aplicación de técnicas como K-Means para clasificar instancias en actividades de la vida diaria, seguida por el entrenamiento de redes neuronales LSTM para clasificar nuevas instancias a partir de las anteriores, demuestra la capacidad de extraer y aplicar conocimiento a partir de los datos, cumpliendo con el tercer objetivo parcial. Aparte, la combinación de técnicas como DBSCAN para inferir la distribución circular de los datos de una misma actividad a lo largo del día permite exponer de manera simple aquellos valores que son anómalos con respecto al resto. Con estas instancias anómalas, un cuidador podrá evaluar los comportamientos para poder identificar situaciones subyacentes que puedan ser preocupantes.
- **Realizar una evaluación empírica, mediante la aproximación del escenario a un entorno real:** La implementación y prueba de la API y los modelos de aprendizaje automático en mi propio hogar, habitado por una persona mayor, representa una aproximación al escenario en un entorno real. Esta evaluación empírica, desde el mes de enero hasta el mes de mayo, ha permitido validar la eficacia del marco de adquisición de datos y también asegurar que el sistema es capaz de operar de manera no intrusiva y respetuosa con la privacidad de los individuos, lo que confirma la consecución del cuarto objetivo parcial.

### 6.3. TRABAJO FUTURO

Aunque el proyecto ya ha dado sus primeros resultados y cumple los primeros objetivos, aún quedan aspectos clave del proyecto a realizar. En este apartado se aportan ideas que serán añadidas en el futuro y ayudarán a crear una herramienta útil en el mundo real por cuidadores y personal médico.

1. **Creación de un *dataset* público propio con los datos recogidos por nuestros enchufes:** Con toda la información recogida de manera propia en varios hogares, crear un *dataset* propio que pueda ser público y accesible, igual que el usado para entrenar el modelo expuesto en este TFM, para que otros investigadores puedan trabajar con ellos.
2. **Creación de una interfaz visual para la información:** Para una mejor visualización de la información extraída a través de los enchufes y almacenada con el formato correspondiente en la base de datos de MongoDB, se podría crear una aplicación web que muestre aquella información relevante para cada enchufe/actividad o incluso una aplicación de utilidad para profesionales y facultativos, en caso de encontrar comportamientos anómalos.
3. **Integración con dispositivos de asistencia por voz:** Se están llevando a cabo proyectos con dispositivos asistentes para hogares de personas mayores y, por tanto, se pretende integrar ambos proyectos con el objetivo de poder complementar el análisis en la detección de ADL y comportamientos anormales, enriqueciendo los resultados y de cara a la toma de decisiones por profesionales

### 6.4. JUSTIFICACIÓN DE COMPETENCIAS ADQUIRIDAS

Siguiendo la normativa vigente en la ESI, se justifican las competencias específicas adquiridas del Máster de Ingeniería Informática tras el desarrollo de este proyecto como TFM y la consecución de los objetivos general y parciales en la Tabla 6.1.

### 6.5. AGRADECIMIENTOS

El presente Trabajo Fin de Master forma parte de las contribuciones derivadas del proyecto nacional financiado "Transformación inteligente del proceso de cuidados basados en actividades de la vida diaria (ADL) y su intervención con soluciones digitalesRef. TED2021-130296A-100. Programa Nacional de Transición Ecológica y Transición Digital (MCIN/AEI/10.13039/501100011033) y NextGenerationEU/PRTR.

**Tabla 6.1:** Competiciones del Máster de Ingeniería Informática

<b>Competencia</b>	<b>Justificación</b>
CE1 - Capacidad para la integración de tecnologías, aplicaciones, servicios y sistemas propios de la Ingeniería Informática, con carácter generalista, y en contextos más amplios y multidisciplinares.	Se justifica esta competencia al integrar un sistema físico para la recolección de información con un sistema de inteligencia artificial para la extracción de conocimiento a través de la información recogida.
CE3 - Capacidad para la dirección de proyectos de investigación, desarrollo e innovación, en empresas y centros tecnológicos, con garantía de la seguridad para las personas y bienes, la calidad final de los productos y su homologación.	Al integrarme en un equipo de investigación, participar en un proyecto de investigación financiado y adquirir los conocimientos necesarios para desempeñar estas tareas que justifican la competencia.
CE4 - Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.	Al diseñar y definir un <i>framework</i> , y luego implantar y operar la fuente primaria de recolección de información que va desde los enchufes hasta la base de datos del sistema, se cumple esta competencia.
CE8 - Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.	Se demuestra al construir una arquitectura que va desde la recogida de información a través de enchufes, hasta su análisis, una vez procesados y almacenados, se demuestra la construcción completa de un sistema de información que cumple las necesidades planteadas.
CE12 - Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.	Al desarrollar un programa que emplea técnicas de aprendizaje automático para la detección precisa de actividades de la vida diaria a través del análisis de datos recogidos de enchufes inteligentes, se cumple esta competencia.

# Bibliografía

---

- [1] Jesús Salido. Curso:  $\LaTeX$  esencial para preparación de tfg, tesis y otros documentos académicos. URL: [http://visilab.etsii.uclm.es/?page\\_id=1468](http://visilab.etsii.uclm.es/?page_id=1468), 2010. Último acceso: sep. 2021.
- [2] Daniel N Mendelson and William B Schwartz. The effects of aging and population growth on health care costs. *Health affairs*, 12(1):119–125, 1993.
- [3] Instituto Nacional de Estadística. Proyecciones de población 2022-2072. URL: [https://ine.es/prensa/pp\\_2022\\_2072.pdf](https://ine.es/prensa/pp_2022_2072.pdf), Oct 2022.
- [4] Instituto Nacional de Estadística. Encuesta continua de hogares. URL: [https://ine.es/prensa/ech\\_2020.pdf](https://ine.es/prensa/ech_2020.pdf), Apr 2021.
- [5] Sidney Katz, Amasa B. Ford, Roland W. Moskowitz, Beverly A. Jackson, and Marjorie W. Jaffe. Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *JAMA*, 185(12):914–919, 09 1963.
- [6] William D Spector, Sidney Katz, John B Murphy, and John P Fulton. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *Journal of chronic diseases*, 40(6):481–489, 1987.
- [7] Ejaz Ahmed, Ibrar Yaqoob, Abdullah Gani, Muhammad Imran, and Mohsen Guizani. Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges. *IEEE Wireless Communications*, 23(5):10–16, 2016.
- [8] Alireza Souri, Aseel Hussien, Mahdi Hoseyninezhad, and Monire Norouzi. A systematic review of iot communication strategies for an efficient smart environment. *Transactions on Emerging Telecommunications Technologies*, 33(3):e3736, 2022.
- [9] Lili Liu, Eleni Stroulia, Ioanis Nikolaidis, Antonio Miguel-Cruz, and Adriana Rios Rincon. Smart homes and home health monitoring technologies for older adults: A systematic review. *International journal of medical informatics*, 91:44–59, 2016.
- [10] Florence I Mahoney and Dorothea W Barthel. Functional evaluation: the barthel index: a simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. *Maryland state medical journal*, 1965.
- [11] M Powell Lawton and Elaine M Brody. Assessment of older people: self-maintaining and instrumental activities of daily living. *The gerontologist*, 1969.
- [12] Margaret Holbrook and Clive E Skilbeck. An activities index for use with stroke patients. *Age and ageing*, 12(2):166–170, 1983.
- [13] George Demiris, Marilyn J Rantz, Myra A Aud, Karen D Marek, Harry W Tyrer, Marjorie Skubic, and Ali A Hussam. Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study. *Medical informatics and the Internet in medicine*, 29(2):87–94, 2004.
- [14] Lode Vuegen, Bert Van Den Broeck, Peter Karsmakers, Bart Vanrumste, et al. Automatic monitoring of activities of daily living based on real-life acoustic sensor data: A preliminary study. In *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies*, pages 113–118, 2013.

- [15] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 51–60, 2009.
- [16] Jung-Yoon Kim, Na Liu, Hwee-Xian Tan, and Chao-Hsien Chu. Unobtrusive monitoring to detect depression for elderly with chronic illnesses. *IEEE Sensors Journal*, 17(17):5694–5704, 2017.
- [17] Sara Casaccia, Gian Marco Revel, and Lorenzo Scalise. Ai-based sensor network for adls monitoring on ageing people during covid-19 epidemic. In *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT)*, pages 290–294. IEEE, 2021.
- [18] Guillaume Sacco, Véronique Joumier, Nelly Darmon, Arnaud Dechamps, Alexandre Derreumaux, Ji-Hyun Lee, Julie Piano, Nathalie Bordone, Alexandra Konig, Bernard Teboul, et al. Detection of activities of daily living impairment in alzheimer’s disease and mild cognitive impairment using information and communication technology. *Clinical interventions in aging*, pages 539–549, 2012.
- [19] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.
- [20] Ioulietta Lazarou, Anastasios Karakostas, Thanos G Stavropoulos, Theodoros Tsompanidis, Georgios Meditskos, Ioannis Kompatsiaris, and Magda Tsolaki. A novel and intelligent home monitoring system for care support of elders with cognitive impairment. *Journal of Alzheimer’s Disease*, 54(4):1561–1591, 2016.
- [21] José M Alcalá, Jesús Ureña, Álvaro Hernández, and David Gualda. Assessing human activity in elderly people using non-intrusive load monitoring. *Sensors*, 17(2):351, 2017.
- [22] Michael A Devlin and Barry P Hayes. Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data. *IEEE transactions on consumer electronics*, 65(3):339–348, 2019.
- [23] Gyubaek Kim and Sanghyun Park. Pre-trained non-intrusive load monitoring model for recognizing activity of daily living. *Applied Intelligence*, 53(9):10937–10955, 2023.
- [24] Hao Song, Georgios Kalogridis, and Zhong Fan. Short paper: Time-dependent power load disaggregation with applications to daily activity monitoring. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pages 183–184. IEEE, 2014.
- [25] Xin Hong, Chris D Nugent, Dewar D Finlay, and Maurice Mulvenna. Homeadl for adaptive adl monitoring within smart homes. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3324–3327. IEEE, 2008.
- [26] Haider Mshali, Tayeb Lemlouma, Maria Moloney, and Damien Magoni. A survey on health monitoring systems for health smart homes. *International Journal of Industrial Ergonomics*, 66:26–56, 2018.
- [27] Anders Krogh, Michael Brown, I Saira Mian, Kimmen Sjölander, and David Haussler. Hidden markov models in computational biology: Applications to protein modeling. *Journal of molecular biology*, 235(5):1501–1531, 1994.
- [28] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29:131–163, 1997.
- [29] Hanna M Wallach et al. Conditional random fields: An introduction. *University of Pennsylvania CIS Technical Report MS-CIS-04-21*, 24:33–42, 2004.
- [30] Douglas A Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.
- [31] Asa Ben-Hur, David Horn, Hava T Siegelmann, and Vladimir Vapnik. Support vector clustering. *Journal of machine learning research*, 2(Dec):125–137, 2001.

- [32] Jinming Zou, Yi Han, and Sung-Sau So. Overview of artificial neural networks. *Artificial neural networks: methods and applications*, pages 14–22, 2009.
- [33] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [34] Soumi Ghosh and Sanjay Kumar Dubey. Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4), 2013.
- [35] Prateek Srivastava and WC Wong. Hierarchical human activity recognition using gmm. In *AMBIENT 2012: The Second International Conference on Ambient Computing, Applications, Services and Technologies*, pages 32–37, 2012.
- [36] Wee-Hong Ong, Takafumi Koseki, and Leon Palafox. An unsupervised approach for human activity detection and recognition. *Int. J. Simul. Syst. Sci. Technol*, 14(5):42–49, 2013.
- [37] Pietro Cottone, Giuseppe Lo Re, Gabriele Maida, and Marco Morana. Motion sensors for activity recognition in an ambient-intelligence scenario. In *2013 IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops)*, pages 646–651. IEEE, 2013.
- [38] K Nanthini and R Manjula Devi. Adaptive fuzzy c-means for human activity recognition. In *International Conference on Information Communication and Embedded Systems (ICICES2014)*, pages 1–5. IEEE, 2014.
- [39] Bo-Chao Cheng, Yi-An Tsai, Guo-Tan Liao, and Eui-Seok Byeon. Hmm machine learning and inference for activities of daily living recognition. *The Journal of Supercomputing*, 54:29–42, 2010.
- [40] Svebor Karaman, Jenny Benois-Pineau, Rémi Mégret, Julien Piquier, Yann Gaëstel, and Jean-François Dartigues. Activities of daily living indexing by hierarchical hmm for dementia diagnostics. In *2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 79–84. IEEE, 2011.
- [41] Md Kamrul Hasan, Husne Ara Rubaiyeat, Yong-Koo Lee, and Sungyoung Lee. A reconfigurable hmm for activity recognition. In *2008 10th International Conference on Advanced Communication Technology*, volume 1, pages 843–846. IEEE, 2008.
- [42] Kai Zhan, Steven Faux, and Fabio Ramos. Multi-scale conditional random fields for first-person activity recognition. In *2014 IEEE international conference on pervasive computing and communications (PerCom)*, pages 51–59. IEEE, 2014.
- [43] Megha Agarwal and Peter Flach. Activity recognition using conditional random field. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*, pages 1–8, 2015.
- [44] Ehsan Nazerfard, Barnan Das, Lawrence B Holder, and Diane J Cook. Conditional random fields for activity recognition in smart environments. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 282–286, 2010.
- [45] Ehsan Nazerfard and Diane J Cook. Using bayesian networks for daily activity prediction. In *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.
- [46] Samer A Mohamed and Uriel Martinez-Hernandez. A light-weight artificial neural network for recognition of activities of daily living. *Sensors*, 23(13):5854, 2023.
- [47] Anthony Fleury, Norbert Noury, and Michel Vacher. Supervised classification of activities of daily living in health smart homes using svm. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6099–6102. IEEE, 2009.
- [48] Nicholas Melo and Jaeryoung Lee. Environment aware adl recognition system based on decision tree and activity frame. *Paladyn, Journal of Behavioral Robotics*, 9(1):155–167, 2018.

- [49] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [50] Salisu Wada Yahaya, Ahmad Lotfi, and Mufti Mahmud. Detecting anomaly and its sources in activities of daily living. *SN Computer Science*, 2(1):14, 2021.
- [51] Sherenaz W Al-Haj Baddar, Alessio Merlo, Mauro Migliardi, et al. Anomaly detection in computer networks: A state-of-the-art review. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 5(4):29–64, 2014.
- [52] Chhavi Dhiman and Dinesh Kumar Vishwakarma. A review of state-of-the-art techniques for abnormal human activity recognition. *Engineering Applications of Artificial Intelligence*, 77:21–45, 2019.
- [53] Nian Chi Tay, Tee Connie, Thian Song Ong, Andrew Beng Jin Teoh, and Pin Shen Teh. A review of abnormal behavior detection in activities of daily living. *IEEE Access*, 11:5069–5088, 2023.
- [54] Sawsan Mousa Mahmoud. *Identification and prediction of abnormal behaviour activities of daily living in intelligent environments*. Nottingham Trent University (United Kingdom), 2012.
- [55] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.
- [56] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [57] Baihong Jin, Yuxin Chen, Dan Li, Kameshwar Poolla, and Alberto Sangiovanni-Vincentelli. A one-class support vector machine calibration method for time series change point detection. In *2019 IEEE International conference on prognostics and health management (ICPHM)*, pages 1–5. IEEE, 2019.
- [58] Salisu Wada Yahaya, Caroline Langensiepen, and Ahmad Lotfi. Anomaly detection in activities of daily living using one-class support vector machine. In *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK*, pages 362–371. Springer, 2019.
- [59] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):1–39, 2012.
- [60] Zhangyu Cheng, Chengming Zou, and Jianwei Dong. Outlier detection using isolation forest and local outlier factor. In *Proceedings of the conference on research in adaptive and convergent systems*, pages 161–168, 2019.
- [61] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.
- [62] Damla Arifoglu and Abdelhamid Bouchachia. Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Computer Science*, 110:86–93, 2017.
- [63] Matias Garcia-Constantino, Alexandros Konios, Idongesit Ekerete, S-RG Christopoulos, Colin Shewell, Chris Nugent, and Gareth Morrison. Probabilistic analysis of abnormal behaviour detection in activities of daily living. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 461–466. IEEE, 2019.
- [64] Peter J Rousseeuw and Mia Hubert. Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(2):e1236, 2018.
- [65] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

- [66] Óscar Belmonte-Fernández, Antonio Caballer-Miedes, Eris Chinellato, Raúl Montoliu, Emilio Sansano-Sansano, and Rubén García-Vidal. Anomaly detection in activities of daily living with linear drift. *Cognitive Computation*, 12(6):1233–1251, 2020.
- [67] Toni Stokes Jones and Rita C Richey. Rapid prototyping methodology in action: A developmental study. *Educational Technology Research and Development*, 48(2):63–80, 2000.
- [68] Dimitri Belli, Paolo Barsocchi, and Filippo Palumbo. Connectivity standards alliance matter: State of the art and opportunities. *Internet of Things*, page 101005, 2023.
- [69] Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2(1):1–14, 2015.
- [70] Changho Shin, Eunjung Lee, Jeongyun Han, Jaeryun Yim, Wonjong Rhee, and Hyoseop Lee. The enertalk dataset, 15 hz electricity consumption data from 22 houses in korea. *Scientific data*, 6(1):193, 2019.
- [71] J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD)*, San Diego, CA, volume 25, pages 59–62. Citeseer, 2011.
- [72] David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1):1–12, 2017.
- [73] Jinlei Zhang, Feng Chen, and Qing Shen. Cluster-based lstm network for short-term passenger flow forecasting in urban rail transit. *IEEE Access*, 7:147653–147671, 2019.
- [74] Lei Liu, Hamed Mosavat-Jahromi, Lin Cai, and David Kidston. Hierarchical agglomerative clustering and lstm-based load prediction for dynamic spectrum allocation. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE, 2021.
- [75] Mohamed Sayah, Djillali Guebli, Noureddine Zerhouni, and Zeina Al Masry. Towards distribution clustering-based deep lstm models for rul prediction. In *2020 Prognostics and Health Management Conference (PHM-Besançon)*, pages 253–256. IEEE, 2020.
- [76] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.
- [77] Adrián Sánchez-Miguel Ortega, Jesús Fontecha Diezma, Iván González Díaz, Luis Cabañero Gómez, and Christopher Nugent. Towards abnormal behaviour detection on elderly people at home through smart plugs and its relationship with activities of daily living. In José Bravo and Gabriel Urzáiz, editors, *Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023)*, pages 118–123, Cham, 2023. Springer Nature Switzerland.
- [78] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.
- [79] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.
- [80] Luis Cabañero, Alejandro Perez-Vereda, Chris Nugent, Ian Cleland, Ramon Hervás, and Iván González. A software tool and a metamodel for digital twins of inhabited smart environments. In *International Conference on Ubiquitous Computing and Ambient Intelligence*, pages 747–759. Springer, 2022.



# **ANEXOS**



## Repositorios con el código del proyecto

El código del proyecto está separado en dos partes. La primera contiene el código en Rust adaptada para el proyecto, disponible en un repositorio en GitHub dentro de la organización del laboratorio MAmI <https://github.com/MAmILab/TAICare-api-extractionSP>.

La segunda contiene el modelo de inteligencia artificial. Para la realización de pruebas, a lo largo del tiempo se utilizó un Notebook en *Google Colab* para tener acceso a sus recursos de GPU debido a problemas con el ordenador de sobremesa del laboratorio <https://colab.research.google.com/drive/1JKU7F5Ldwluto8A0bgfoz-LG0fR6DybT?usp=sharing>. Una vez resueltos estos problemas, se terminó el Notebook y se refactorizó el código, creando sus correspondientes clases y archivos, y se subió a otro repositorio en GitHub dentro de la organización anteriormente mencionada <https://github.com/MAmILab/TAICare-notebook-code>.