

Instituto Politécnico de Viseu

Escola Superior de Tecnologia e Gestão de Viseu

David Batista da Mota

Monitorização de padrões de movimento em idosos  
no domicílio

**Tese de Mestrado**

Sistemas e Tecnologias de Informação para as Organizações

Professor Doutor Carlos Augusto da Silva Cunha

Professor Doutor Rui Pedro Monteiro Amaro Duarte



Abril de 2022



*Por todas as razões este projeto representa uma enorme superação. Ainda assim,  
tal não teria sido possível sem o apoio da Família, Pais Irmãos e Sogros..  
Um especial obrigado às minhas três cúmplices, Catarina, Carolina e Constança...*



# Agradecimentos

Aos meus Orientadores uma palavra amiga, por sempre terem acreditado no projeto, na sua pertinência, na sua atualidade e na minha capacidade de o executar.



# Resumo

O envelhecimento da população é uma tendência confirmada pelos últimos estudos realizados à população, tanto a nível nacional como internacional. Assim, garantir qualidade de vida e uma adequada monitorização, nas mais diversas dimensões dos idosos nas suas habitações, revela-se uma prioridade e, naturalmente, uma área de estudo relevante. A identificação de padrões de movimento e o treino de um algoritmo de *machine learning* para a deteção de *outliers*, representam os objetivos principais do trabalho.

A monitorização de pessoas está frequentemente associada à perda de privacidade e autonomia. De forma a evitar estas situações, propusemo-nos explorar a capacidade de reutilizar soluções de IoT e sensorização existentes na habitação. Pretende-se extrair das soluções de *smart homes*, o máximo de informação possível para, numa primeira fase, estudar os padrões de movimento dos habitantes e, dessa forma, alavancar um conjunto de ações que potenciem o seu bem estar e monitorização. Para atingir os objetivos propostos no projeto, foram implementadas várias abordagens para o estudo de padrões, a análise da distribuição dos registos, o calculo dos percentis e o calculo da probabilidade da ocupação. Implementou-se um algoritmo de *machine learning* para proceder à identificação de *outliers* nos dados extraídos dos sensores da habitação.

**Palavras-Chave:** Geriatria, *Internet of Things* (IoT), Envelhecimento, Sensores, Smart Home, Monitorização, Outliers.



# Abstract

The aging of the population is a trend that has been confirmed by national and international studies. Identifying and studying the movement patterns as well as the training of a machine learning model to identify outliers, were the main goals of this project. Ensuring quality of life and an adequate monitorization of the elderly population in their homes has proven to be a priority and a relevant area of research. The monitorization of people is frequently associated with privacy and autonomy issues. In order to be able to avoid this, existing IoT and *Smart Home Solutions* was explored. The aim was to extract from *Smart Home* solutions as much information as possible so that, movement patterns could be identified and studied and thereby leverage a set of improvements of the monitoring process. To achieve these improvements a machine learning algorithm was implemented, so that, *outliers* could be identified in the data that was extracted from the sensors.

**Keywords:** Geriatrics, IoT, Sensors, Smart Home, Monitoring, Outliers, elderly population



# Índice

<b>Lista de Figuras</b>	<b>vii</b>
<b>Código fonte</b>	<b>ix</b>
<b>Lista de Acrónimos</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Contextualização . . . . .	2
1.3 Definição do Problema . . . . .	4
1.4 Objetivos . . . . .	5
1.5 Resultados esperados . . . . .	6
1.6 Plano de Trabalhos . . . . .	6
1.7 Estrutura da Tese . . . . .	7
<b>2 Estado da Arte</b>	<b>9</b>
<b>3 Tecnologias</b>	<b>15</b>
3.1 <i>Internet of Things</i> . . . . .	15
3.2 <i>Smart Home</i> . . . . .	16
3.3 <i>Smart Home Privacy</i> . . . . .	17
3.4 Sensores . . . . .	18
3.5 <i>Home Automation</i> . . . . .	19
3.6 Python . . . . .	21
3.7 Matplot lib . . . . .	22
<b>4 Metodologia</b>	<b>23</b>
4.1 Nota Introdutória . . . . .	23
4.2 Contexto observado . . . . .	24
4.3 Acesso aos dados . . . . .	25
4.4 Extração e pré-processamento de dados . . . . .	26
4.5 Dataset . . . . .	27
4.6 Detecção de Anomalias . . . . .	28

---

<b>5</b>	<b>Implementação</b>	<b>31</b>
5.1	Arquitetura . . . . .	31
5.2	Localização dos sensores . . . . .	32
5.3	Recolha de dados . . . . .	35
5.4	Integração dos dados . . . . .	35
5.5	Caracterização do <i>dataset</i> . . . . .	37
5.6	Pré-processamento de dados . . . . .	38
5.6.1	Eliminação de períodos incompletos . . . . .	38
5.6.2	Agrupamento de dados . . . . .	38
5.6.3	Formatação de <i>features</i> . . . . .	39
5.7	Reconhecimento de padrões . . . . .	40
5.8	Algoritmo de <i>Machine Learning</i> . . . . .	41
<b>6</b>	<b>Análise de Resultados</b>	<b>43</b>
6.1	Identificação de padrões . . . . .	43
6.1.1	Acumulado de registos . . . . .	43
6.1.2	Distribuição Temporal das Presenças . . . . .	45
6.1.3	Probabilidade de Ocupação . . . . .	47
6.2	Deteção de <i>outliers</i> . . . . .	50
6.3	Discussão de resultados . . . . .	52
<b>7</b>	<b>Conclusão e trabalho futuro</b>	<b>53</b>
	<b>Referências</b>	<b>55</b>

# Lista de Figuras

3.1	Dispositivos - Informação fornecida vs Privacidade [Debes et al., 2016]	17
3.2	Sensor <i>Passive Infra Red sensor</i> (PIR) [Aqara, 2021]	18
3.3	Ângulo máximo do sensor PIR [Aqara, 2021]	18
3.4	Smart Homes [Sovacool and Furszyfer Del Rio, 2020]	20
4.1	Acesso aos dados dos sensores via OpenHab	25
4.2	OpenHab <i>components</i> [Parocha and Macabebe, 2019]	26
4.3	Carregamento dos dados para MongoDB	27
5.1	Arquitetura da solução de aquisição e integração de dados	32
5.2	Quarto (Sensor 1)	33
5.3	Casa de banho (Sensor 2)	34
5.4	Sala de estar (Sensor 3)	34
5.5	Cozinha (Sensor 4)	34
5.6	Portal OpenHab®	35
5.7	Excerto do dataset original	37
5.8	Excerto do <i>Dataset</i>	40
6.1	Total de ocorrências por hora no compartimento <i>Cozinha</i>	44
6.2	Total de ocorrências por hora no compartimento <i>Sala</i>	44
6.3	Total de ocorrências por hora no compartimento <i>Casa de banho</i>	45
6.4	Total de ocorrências por hora no compartimento <i>Quarto</i>	45
6.5	Disparos no compartimento <i>Cozinha</i> . Percentil 95 e 5	46
6.6	Disparos no compartimento <i>Sala</i> . Percentil 95 e 5	46
6.7	Disparos no compartimento <i>Casa de Banho</i> . Percentil 95 e 5	47
6.8	Disparos no compartimento <i>Quarto</i> . Percentil 95 e 5	47
6.9	Probabilidade de ocupação da <i>Cozinha</i>	48
6.10	Probabilidade de ocupação da <i>Sala</i>	48
6.11	Probabilidade de ocupação da <i>Casa de Banho</i>	49
6.12	Probabilidade de ocupação do <i>Quarto</i>	49
6.13	Probabilidade de ocupação em todos os compartimentos	49
6.14	<i>Histograma Outliers</i>	50
6.15	Listagem dos registos <i>outliers</i>	51
6.16	Registo "normal"	51

6.17 <i>Outlier</i> detectado . . . . .	51
---	----

# Código fonte

5.1	OpenHab log . . . . .	36
5.2	Registo do evento gerado no processo de integração . . . . .	37



# Lista de Acrónimos

<b>AAL</b>	<i>Ambient Assisted Living</i>
<b>ADL</b>	<i>Activities of Daily Living</i>
<b>API</b>	Aplication Programming Interface
<b>Bluetooth</b>	<i>Short-range Wireless Technology Standard</i>
<b>ETL</b>	<i>Extract Transform Load</i>
<b>IoT</b>	<i>Internet of Things</i>
<b>JSON</b>	<i>Java Script Object Notation</i>
<b>K-Mean</b>	k-means clustering
<b>ML</b>	<i>Machine Learning</i>
<b>NaN</b>	Not a Number
<b>non-SQL</b>	<i>Non Relational Database</i>
<b>OECD</b>	<i>Organization for Economic Co-operation and Development</i>
<b>OMS</b>	Organização Mundial da Saúde
<b>PC</b>	<i>Personal Computer</i>
<b>PIR</b>	<i>Passive Infra Red sensor</i>
<b>RAM</b>	<i>Random-Access Memory</i>
<b>Wifi</b>	<i>IEEE 802.11b Direct Sequence</i>
<b>Zigbee</b>	<i>Zonal Intercommunication Global-standard</i>



# Capítulo 1

## Introdução

Neste capítulo são expostas as motivações para a realização deste trabalho, bem como a relevância do tema no contexto dos ambientes de vida assistida e, por fim, descreve-se a estrutura utilizada na construção do documento.

### 1.1 Motivação

Atualmente é sabido que a tecnologia está presente em grande parte do dia a dia da maioria das populações, assumindo um papel muitas vezes fundamental, vital, e do qual já praticamente não se consegue abdicar. Sintetizar as áreas da sociedade onde a tecnologia está presente, ou os diferentes contextos onde possa ser utilizada, representa certamente uma tarefa árdua e extensa. Pensar que a tecnologia que nos rodeia esconde um potencial enorme no que diz respeito ao auxílio no estudo de problemas de saúde é, no mínimo, desafiante. Estudar e entender a melhor forma de tirar proveito da tecnologia sem a necessidade de a recriar ou tornar mais complexa a sua utilização, constituiu também um fator de grande motivação.

Dado o contexto de população envelhecida e a proliferação de soluções tecnológicas ao nosso redor, torna-se fundamental tirar partido da tecnologia existente para a colocar ao serviço de outros contextos, nomeadamente na monitorização e estudo dos padrões de movimentação de idosos no domicílio. A monitorização inteligente e não intrusiva dos idosos, recorrendo ao tratamento de dados para o treino de modelos de *Machine Learning* (ML) não supervisionados, sustenta cientificamente a realização deste trabalho. Ainda que se trate de um protótipo, esta solução poderá também

servir de ferramenta para auxiliar os atuais prestadores de cuidados. Não menos relevante, o potencial papel que a solução possa desempenhar no fornecimento de informação para os profissionais de saúde que pretendam estudar e avaliar o dia a dia do idoso.

## 1.2 Contextualização

Segundo a *Organization for Economic Co-operation and Development* (OECD) a população idosa irá aumentar na maioria dos países e, nesse sentido, Portugal mantém o quarto lugar na lista de países com maior taxa de população idosa, tanto para a faixa etária com mais de 65 anos como para a dos 80 anos ou mais [OECD, 2021]. Estes indicadores revelam um acentuado envelhecimento da população.

Com a entrada na fase geriátrica da vida (período a cima dos 65 anos de idade), o idoso confronta-se com um conjunto de desafios tais como a manutenção da sua qualidade de vida nas mais diferentes dimensões. Inevitavelmente, o ser humano enfrenta um decréscimo das suas capacidades físicas e, muitas vezes, também cognitivas. Com o avançar da idade, podem surgir várias patologias características deste período em que, com maior ou menor gravidade, a autonomia do idoso fica comprometida, o que pode levar a perda de capacidade na realização das suas tarefas diárias [Tornero-Quinones et al., 2020], o que poderá afetar a autonomia do idoso. Por outro lado, o envelhecimento não é algo que dependa apenas de fatores biológicos, mas claramente também, de fatores psicológicos, sociais e ecológicos. É igualmente expectável que seja observada uma capacidade de adaptação e transformação para esta nova realidade. Não só se observa o envelhecimento da população como se confirma a tendência da mesma para querer viver com autonomia e de forma confortável na sua habitação tanto tempo quanto possível [Amián et al., 2021].

Em Portugal, parte da população opta por envelhecer na sua própria habitação, sendo cada vez mais diferenciada e com maior longevidade. A existência de várias soluções públicas e privadas, nomeadamente soluções de apoio e prestação de cuidados ao domicílio reforçam a possibilidade e a manutenção da autonomia e inclusão social dos idosos [Fonseca, 2020]. Desta forma, e com o crescimento da população idosa previsto anteriormente, projetam-se cenários de manutenção da atividade social e da autonomia dos idosos. Ainda assim, o apoio e monitorização do bem-estar desta população revela-se de enorme importância e responsabilidade. Em Portugal, os sistemas de monitorização de idosos, seja pela via da visita física, seja pela instalação de soluções de monitorização, são já uma realidade [Ageing@Coimbra, 2020]. Existem também iniciativas dinamizadas pelos agentes do poder local que visam acompanhar o envelhecimento em casa proporcionando a melhor qualidade de vida ao idoso [ANCS, 2020].

A autonomia do idoso, nomeadamente na execução das tarefas do dia a dia, pode ser comprometida pelo surgimento de patologias, nomeadamente a demência. A demência representa um conjunto de sintomas associados a doenças que provocam a perda de capacidades cognitivas e motoras nas pessoas [Larson, 2021], tipicamente associadas a perda de memória, capacidade intelectual e de raciocínio, entre outras. O reconhecimento de alterações cognitivas e ou motoras, raramente é feito pelo próprio. Na maioria dos casos é a família ou cuidadores que partilham com o Médico as alterações observadas e, mesmo assim, por vezes de uma forma tardia associando estas alterações de comportamento ao processo de envelhecimento, levando a um diagnóstico tardio da patologia [Larson, 2021].

Em todo o mundo, segundo a Organização Mundial da Saúde (OMS), estima-se que possam existir cerca de 47 milhões de pessoas com demência, sendo provável que este valor possa vir a triplicar até 2050. A doença de Alzheimer representa entre 60% a 70% dos casos de demências. Em Portugal estimam-se que até 2050 tenhamos 374 mil casos, afetando assim 3,82% da população portuguesa [alzheimerportugal, 2020].

Em muitas situações a monitorização do declínio da capacidade de realização das atividades, nomeadamente na identificação de situações de alteração à normal realização as atividades diárias [Cipriani et al., 2020] assume um importante papel precisamente na identificação de situações de risco para o idoso.

A monitorização de idosos poderá passar pela utilização dos atuais sistemas de segurança no domicílio, ou de domótica, que geram dados em grande quantidade. Estes dados podem estar a ser ignorados do ponto de vista analítico, contudo, se tratados, podem representar valor para o estudo dos hábitos das pessoas nas suas próprias habitações. Ao incidir o foco nos idosos na sua habitação, estes dispositivos podem gerar informação extremamente valiosa, não só para a monitorização da sua segurança e do seu bem estar, mas até para possibilitar o estudo e a identificação de alterações dos padrões de vida que, por sua vez, podem estar associados a patologias tais como as demências. Assim, a monitorização das atividades do dia a dia do idoso, pode ajudar na avaliação da sua autonomia, visto que esta só lhe é reconhecida enquanto as consegue realizar.

O termo *Activities of Daily Living* (ADL) caracteriza um conjunto de atividades identificadas como essenciais no dia a dia do ser humano e que deverão ser executadas de forma independente pelo indivíduo. A incapacidade de cumprir estas atividades pode indiciar situações de insegurança e até a redução da qualidade de vida [Edemekong et al., 2021].

As pessoas idosas tendem a prolongar a permanência na sua própria habitação tanto quando possível, sendo esta permanência vista muitas vezes como sinal de capacidade de manutenção das suas atividades diárias. Ainda assim, infelizmente, a perda de autonomia é raramente valorizada pelo próprio e por sua vez identificada. Muitas vezes, esta perda acaba por ser identificada pelos profissionais de

saúde em situações de auxílio e ida às urgências. Outra forma de identificar estas situações passa por avaliações através da realização de questionários clínicos [Kwon et al., 2021] e da observação externa por parte de familiares ou próximos. Numa tentativa de mitigar estas situações e a identificação precoce das perdas de autonomia dos idosos, estuda-se a utilidade dos ambientes *Ambient Assisted Living* (AAL) para monitorização e alerta.

Os ambientes AAL têm como principal função auxiliar o indivíduo no seu dia a dia melhorando a sua qualidade de vida. A manutenção desta pode passar por garantir a atividade social do idoso [Pinto et al., 2019], o auxílio na realização de tarefas [Kearney et al., 2018] ou até na monitorização do seu dia a dia [Zambrano-Montenegro et al., 2019, Denecke, 2021, Maskeliūnas et al., 2019].

Ambientes de AAL incorporam não apenas conhecimentos técnicos ou relacionados com os sensores e *software* mas também um conjunto de outros domínios da ciência que partilham o mesmo objetivo, melhorar a qualidade de vida das pessoas. Alguns exemplos de aplicabilidade de soluções de AAL são a deteção de emergências [Sandeepa et al., 2020] (utilização de um sistema de deteção de falha cardíaca recorrendo a sensores *wearables*) e apoio no diagnóstico e acompanhamento de pessoas com doenças [Ghayvat et al., 2019] (recurso a um conjunto de diferentes sensores para identificar ADL e eventuais alterações ao normal comportamento).

A criação de ambientes de AAL hoje em dia está mais facilitada dada a grande disponibilidade de diferentes sensores e plataformas pré-configuradas e disponíveis no mercado (assistentes virtuais ou sensores, interruptores e lâmpadas inteligentes), que podem desde logo auxiliar o utilizador em diversas tarefas do seu dia a dia e, dessa forma melhorar a sua qualidade de vida.

### 1.3 Definição do Problema

O envelhecimento da população e a opção por viver o período geriátrico na própria habitação motivam a criação de soluções ao nível do acompanhamento e da monitorização da sua segurança e bem estar. Esta informação é relevante não só para o próprio como para os clínicos e familiares.

Ao nível do acompanhamento, a gradual perda de capacidades cognitivas e motoras, pode desencadear situações de risco. Estas situações podem caracterizar-se por quedas, a não realização das atividades diárias e sedentarismo, entre outras, que são frequentemente identificadas pelos cuidadores (sejam família ou empresas contratadas para o efeito ou até, quando existem, vizinhos). Infelizmente, muitas vezes estas situações não são detetadas em tempo útil.

Por outro lado, a monitorização de idosos pode ser feita de várias formas, recorrendo a familiares, a outros agentes locais ou mesmo, em ultima instância, à institucionalização dos idosos. A maioria das soluções de prestação de cuidados,

tais como lares e centros de dia, pode representar outros desafios, como capacidades financeiras para suportar a contratação de profissionais, eventuais alterações de hábitos, invasão de privacidade e até abandono da habitação. Algumas destas mudanças nem sempre agradam ao idoso ou representam uma manutenção ou melhoria da qualidade de vida. A opção por sistemas de monitorização não intrusivos de baixo custo pode representar uma alternativa às soluções anteriormente partilhadas. A solução deve ser autónoma e de fácil instalação, não obrigando o idoso a qualquer tipo de interação no dia a dia. Para mitigar os erros e riscos associados aos *wearables*, a solução não deve obrigar o idoso a vestir ou transportar consigo diferentes sensores.

Dada a existência de soluções de *Smart Home* que contam com sensores de movimento espalhados por toda a habitação, surge o desafio de aceder e explorar os dados gerados por estes sensores de forma a avaliar a sua pertinência para o estudo de padrões de movimento, bem como de respetivas alterações.

## 1.4 Objetivos

Considerando os desafios apresentados nas secções anteriores, este trabalho tem como objetivo principal a criação de uma solução de monitorização da habitação através de sensores *Passive Infra Red sensor (PIR)*, que permitam a inferência de padrões de movimento para verificação de situações anómalas num ambiente de AAL.

Mais concretamente, este objetivo é decomposto em vários sub-objetivos que o completam:

- *(O1) - Instalação de uma rede de sensores:* Numa habitação é instalada uma rede de sensores de forma a poder monitorizar 24 horas por dia, todas as movimentações nas divisões que compõem a habitação.
- *(O2) - Pré-processamento dos dados:* Após a instalação da rede de sensores, o *hub* recebe os dados fornecidos dos sensores. Estes dados são guardados e tratados de forma segura, para posterior estudo.
- *(O3) - Processamento dos dados:* Os dados tratados e transformados possibilitam o estudo e identificação de padrões de movimento associados às rotinas dos habitantes.
- *(O4) - Treino de modelo de ML:* Com os dados existentes, pretende-se treinar um modelo de ML que permite criar condições para a identificação de situações anómalas que venham a ser verificadas.

## 1.5 Resultados esperados

De acordo com os objetivos identificados na secção anterior, esperam-se obter os seguintes resultados:

- *Deteção de padrões de movimento*: Atendendo a que os dados recolhidos se tratam de dados temporais, deverão permitir a análise e identificação de padrões de movimento. Para serem reconhecidos padrões, os dados devem revelar consistência na sua distribuição ao longo do dia nos diferentes dias da semana. Os padrões que sejam identificados devem permitir a identificação de diferentes rotinas nas pessoas que habitam um espaço. Estes resultados decorrem dos objetivos *O1 - O3* enunciados na secção anterior.
- *Deteção de alterações aos padrões de movimento nas rotinas diárias*: É expectável que a quantidade de dados recolhidos seja suficiente para se proceder ao treino de modelos de ML. Por fim, o sistema deverá conseguir identificar situações onde se verifiquem alterações ao padrão de comportamento tipificado como normal de pessoas numa habitação. Estes resultados estão relacionados com o objetivo *O4*.

## 1.6 Plano de Trabalhos

De forma a atingir os objetivos enunciados anteriormente, com foco nos resultados esperados, o desenvolvimento deste trabalho contou com as seguintes macro etapas:

- Autorização prévia dos monitorizados para a recolha de dados: Antes da realização do trabalho, foi requerido aos participantes uma autorização para recolha de dados de monitorização, uma vez que, sem esta, seria impossível a realização do trabalho. De acordo com o Regulamento Geral de Proteção de Dados (RGPD) [Magalhães and Pereira, 2020], o tratamento dos dados recolhidos neste trabalho baseia-se no consentimento livre, informado e explícito, assegurando a conformidade de transparência na recolha e na defesa de interesses do titular dos dados. Para além disso, o titular dos dados tem o direito de retirar o consentimento a qualquer momento.
- estudo e avaliação de trabalhos relacionados com conhecimento relevantes para a sustentação e realização desta dissertação: nesta fase, será realizada uma revisão bibliográfica com especial foco em trabalhos e tecnologias relacionadas com o tema deste trabalho.
- seleção, configuração e instalação do hardware necessário para a uma adequada recolha de dados (unidade de processamento, sensores e hub)

- desenvolvimento dos mecanismos de extração e tratamento dos dados bem como dos mecanismos de análise dos dados recolhidos
- configuração e treino do modelo de deteção de *outliers*
- análise dos resultados e conclusões finais

Este projeto contou inicialmente com a instalação e configuração dos sensores, sendo posteriormente necessário aguardar pela geração e recolha dos dados. Os dados foram recolhidos durante um período de sete meses.

## 1.7 Estrutura da Tese

Esta tese está organizada e estruturada da seguinte forma: o Capítulo 1, Introdução, fornece ao utilizador as principais motivações e contextualização do tema na atualidade bem como a definição dos objetivos e resultados esperados; o Capítulo 2, Estado da Arte, foca-se em fornecer ao leitor um melhor entendimento sobre as tecnologias relevantes para o projeto, bem como os cenários de investigação e implementação relacionados com o tema; o Capítulo 3, Tecnologias, são descritas as tecnologias envolvidas na realização do projeto; o Capítulo 4, Metodologia, descreve a arquitetura implementada; o Capítulo 5, Implementação, concretiza, na prática, a execução do projeto; o Capítulo 6, Análise de Resultados, apresenta a análise e discussão dos resultados obtidos; o Capítulo 7, Conclusão e Trabalho futuro, apresenta as notas finais, retirando das mesmas conclusões para o futuro. Finalmente, as Referências de todas as fontes consultadas.



## Capítulo 2

# Estado da Arte

A problemática do envelhecimento da população e a conseqüente procura por soluções que auxiliem o idoso na manutenção da sua qualidade de vida, tem desencadeado um conjunto de estudos e experiências [AAL, 2021]. A tecnologia claramente tem um papel relevante na criação de ferramentas que contribuem para um melhor acompanhamento e, conseqüentemente, melhor qualidade de vida em período geriátrico.

As principais necessidades identificadas e que têm despertado maior interesse por parte dos investigadores são a identificação de cenários de queda [Kelly et al., 2019], a gestão nutricional [Casaccia et al., 2019], os problemas de visão [Andò et al., 2015], a depressão [Chen et al., 2018] e o delírio e demência [Kwon et al., 2021]. Os sistemas de monitorização para a saúde podem ainda ser de diferentes tipos: os mais intrusivos, tais como os sensores que ficam acoplados ao corpo do utente [Olmedo-Aguirre et al., 2022] (mais conhecidos por *wearables*), e os não intrusivos ou passivos [Gochoo et al., 2019], tais como sensores de movimento, de temperatura, magnéticos, entre outros [Monteriù et al., 2018].

Os dispositivos de *Internet of Things* (IoT) têm tido uma influência em várias áreas relacionadas com a saúde (os sistemas AAL são um desses exemplos), com um especial foco na monitorização do dia a dia dos pacientes, sendo por isso sistemas com um enorme potencial na identificação de casos de emergência ou deteção de anomalias/alterações no dia a dia das pessoas. Estes sistemas são frequentemente autónomos e dotados de capacidade de comunicação, fornecendo, desta forma, informação aos cuidadores e familiares, bem como aos profissionais de saúde [Sahu et al., 2021].

A monitorização de idosos com potenciais perturbações cognitivas e motoras tem sido alvo de vários estudos. Um deles focou-se na sintetização e análise do tipo de estudos e implementações que têm sido realizadas. No que diz respeito à monitorização das atividades físicas dos idosos, foram utilizadas várias estratégias, isto é, vários tipos de sensores, desde os *wearables* aos sensores de pressão até sensores de movimento para monitorizar a realização das atividades diárias, bem como a monitorização de situações de perigo, tais como as quedas. É ressalvada a importância dos cenários de AAL assim com a falta de aposta em reais implementações destes sistemas em benefício das populações mais idosas [Ganesan et al., 2019].

Em [Rezvani et al., 2021] são recolhidos dados gerados por sensores instalados nos diferentes compartimentos das habitações para identificar e registar situações de agitação. A agitação em pessoas no período geriátrico tem especial interesse para os investigadores uma vez que pode indiciar a presença de doenças relacionadas com a demência. De forma a conseguir identificar melhor as situações de agitação, é sugerida a criação de um modelo de ML auto-supervisionado recorrendo a redes neuronais convulsionais [Li et al., 2021]. A escassez de situações, dados referentes a agitação devidamente classificados, tem um impacto considerável no treino dos modelos (com impacto direto na performance do modelo). Ainda assim, o modelo apresentado e implementado conseguiu evidenciar uma elevada capacidade de deteção de situações de agitação e mostrou ainda ser adaptável a diferentes cenários.

Com o propósito de melhorar o diagnóstico precoce de demências e de forma a complementar um questionário clínico, em Kwon et al. [Kwon et al., 2021] é implementada uma rede de sensores IoT capaz de registar as diferentes atividades do dia a dia de sete pacientes. Desta forma (com a recolha de dados através de sensores) foi possível confrontar os resultados das entrevistas realizadas aos utentes com os resultados da interpretação dos dados dos sensores. Ao contrário das respostas dos questionários, que podem ser subjetivas, a monitorização através de sensores produz resultados reais e representativos da realidade. A estratégia passou por instalar um conjunto de sensores *Zonal Intercommunication Global-standard* (Zigbee) nos diferentes compartimentos frequentados pelos pacientes a serem monitorizados, de forma a recolher os dados gerados para posterior análise. No mesmo projeto foram treinados modelos de ML de forma a tentar identificar alterações à realização das atividades diárias normais uma vez que as alterações à normal realização destas atividades podem representar sintomas relacionados com o aparecimento de algum tipo de demência. Os sensores comunicam com um *hub* que, por sua vez, transmite os dados para uma *cloud* para armazenamento e posterior análise. As diferentes atividades diárias foram catalogadas e podem ser identificáveis através da identificação e catalogação de diferentes sequências de ativação de sensores nos diferentes compartimentos. Foram utilizados métodos estatísticos e mecanismos de ML para identificar e catalogar os diferentes tipos de comportamentos. Foi implementado

um modelo de ML para a classificação de comportamentos e tentativa de diagnóstico precoce de demência. O algoritmo treinado para a identificação de *outliers* foi o *Random Forest* [Breiman and Cutler, 2005]. O trabalho desenvolvido nesta tese está relacionado com este trabalho, no entanto, ao invés de efetuar uma análise aos padrões de movimento, optam por tentar fazer o reconhecimento e classificação as diferentes atividades diárias. O foco está na detecção de alterações de comportamento na realização das tarefas diárias e com isso tentar identificar estádios iniciais de demência.

A avaliação física do idoso foi alvo de estudo em [Mertens et al., 2021], com especial foco na monitorização da sua autonomia. À semelhança do trabalho de [Kwon et al., 2021], em [Mertens et al., 2021] foi realizado um questionário que apresenta fragilidades na avaliação, uma vez que as respostas às perguntas podem não ser assertivas nem representar a real situação do idoso. Assim, com o objetivo de criar um sistema capaz de identificar situações de risco e de as reportar, os autores procederam à implementação de uma rede de sensores que, de forma não supervisionada e não intrusiva, recolhem dados sobre o dia a dia dos utentes e, desta forma, contribuem para a capacidade de distinguir situações normais de situações anormais. Este projeto foi implementado tendo em conta o cenário de monitorização de um indivíduo único por habitação. Neste estudo foi instalado um conjunto de sensores PIR e câmaras nos diferentes compartimentos da habitação com o intuito de monitorizar a existência de presença e a duração da mesma em cada uma das divisões. Os sensores geram informação binária quando acionados (0 representa não presença e 1 representa presença). Os espaços são representados por letras e os dias são representados por sequências de letras. A análise dos desvios aos padrões de normalidade de presença numa divisão é feita através da contabilização do tempo despendido pelo utente em cada uma das divisões ao longo do dia. O tempo despendido é depois comparado com o período semelhante nos diferentes dias recorrendo a métodos estatísticos, nomeadamente através do cálculo da média e do desvio padrão da distância de Hamming [Yang, 2021], de forma a calcular o desvio da normalidade. Cada vez que o valor de desvio é ultrapassado, é registada uma alteração ao padrão de normalidade. Não foram aplicados algoritmos de ML para identificação e comparação de padrões de presença numa divisão. Este trabalho foca-se na identificação e comparação dos padrões de movimento diários, por períodos de cinco dias e um dia, ao invés da monitorização hora a hora (que é usada no trabalho proposto neste documento de tese). Para além disso, são utilizados métodos invasivos (nomeadamente câmaras de vídeo) e não fazer uso de técnicas de ML para a identificação de *outliers*.

Em [Fiorini et al., 2017] são estudados os dados gerados com recurso a sensores PIR com o objetivo de identificar os diferentes padrões de movimento associados a diferentes tipos de habitantes, bem como explorar e aferir qual o volume de dados,

em dias, necessário para garantir que o comportamento de cada um dos indivíduos é padronizado e identificável. Tal comportamento permite determinar se se está na presença de alguém idoso e mais sedentário ou de uma pessoa ativa com um estilo de vida dinâmico. Trata-se então da criação de um mecanismo que não só monitoriza o dia a dia da pessoa como caracteriza o seu estilo de vida. Para além de sensores PIR são também usados sensores magnéticos instalados na porta de entrada da habitação, na porta do frigorífico e em cada um dos compartimentos. Os dados gerados são recolhidos através da utilização da plataforma *Vera Smart Home*, que conta com um gateway com o qual todos os sensores comunicam partilhando dados sobre o seu estado. Os dados são extraídos e armazenados numa base de dados online. Os algoritmos utilizados para estudar os padrões de movimento foram K-Mean [sklearn, 2022] e Self-Organizing Maps (SOM) [Miljkovic, 2017]. São ambos algoritmos que consistem no *clustering* dos dados. Os resultados mostram que, com registos recolhidos durante um período de 55 dias, é possível traçar perfis de diferentes utilizadores e, com isso, prestar serviços de cuidados adequados a cada um dos perfis, potenciando, desta forma, uma prestação de serviços personalizada e ajustada às necessidades de cada agregado.

Em [Yahaya et al., 2021] é identificado o *outlier* nos dados recolhidos bem como a *source* ou *feature* relacionada com o *outlier*, em que Um ponto é identificado como não pertencente a nenhum conjunto de pontos. Ainda assim o facto de estar distante não impede a sua classificação em relação a um eixo específico. Foi utilizado um *dataset* da *Nottingham Trent University*, e do *Center for Advanced Studies in Adaptive System (Washington State University)* [CASAS, 2013]. Foram aplicados vários algoritmos para estudar e modelar os padrões de movimento dos diferentes pacientes. Para efeitos de teste da identificação da *feature* causadora do *outlier* foram introduzidos valores no *dataset* de forma a provocar as alterações. Os resultados mostram que foi possível identificar a *feature* responsável pelo *outlier* com uma precisão de 82%.

A título de conclusão, a literatura focada na monitorização de idosos no domicílio é vasta e muitas abordagens foram e são exploradas para melhor garantir qualidade de vida. O avanço da tecnologia tem possibilitado, igualmente, a proliferação de soluções, sensores e algoritmos que possibilitam a criação de plataformas de grande valor, seja na monitorização das questões de segurança, de saúde, no apoio ao estudo de patologias, entre outras. A recolha de dados e o seu estudo está patente em diferentes estudos, conforme apresentado anteriormente. O trabalho apresentado nesta tese foca-se em três aspetos essenciais: a criação de uma solução de monitorização não intrusiva, de baixo custo e inteligente. A não utilização de *wearables* marca ainda a diferença para outros estudos. Com a garantia do acesso aos dados gerados pelos sensores e com o seu estudo é possível potenciar a criação de uma plataforma de monitorização *as a service*, tanto ao nível da deteção de padrões de movimento,

bem como na detecção de alterações a esses padrões, como será apresentado nos capítulos seguintes.



## Capítulo 3

# Tecnologias

Nas secções que se seguem, são apresentadas as soluções tecnológicas escolhidas e utilizadas neste trabalho. É feita uma contextualização da tecnologia, bem como um mapeamento com o cenário em questão.

### 3.1 *Internet of Things*

Tal como o nome indica, IoT representa um conjunto de dispositivos (*things*) conectados à Internet. Os dispositivos da arquitetura IoT dividem-se em dois tipos, os sensores e os atuadores. Os sensores são ativados maioritariamente com alterações que acontecem no meio físico (sensores de temperatura, de movimento, etc), enquanto que os atuadores são ativados remotamente e executam a tarefa que lhes for atribuída (uma lâmpada inteligente quando liga e desliga). Por vezes, são os próprios sensores que despoletam ações nos atuadores.

A utilização de dispositivos IoT pode auxiliar na simplificação e otimização de processos e sistemas e até melhorar a qualidade de vida das pessoas. Podemos encontrar implementações de redes IoT em diferentes contextos, desde a monitorização ambiental [Doshi et al., 2019], aos edifícios inteligentes [Shah et al., 2019], transportes inteligentes [Zantalis et al., 2019], economia circular [Gligoric et al., 2019], sistemas de AAL [Maskeliunas et al., 2019], entre outros.

Os sensores IoT comunicam frequentemente através de tecnologias como o Zigbee e o Bluetooth [Elhadi et al., 2021]. Por serem sensores de baixo consumo de energia e de reduzidas dimensões, uma vez instalados facilmente passam despercebidos

para o habitante. Os sensores de movimento garantem ainda uma monitorização não intrusiva [Sharma et al., 2021], uma vez que não registam imagem nem som. É importante salientar que o rápido crescimento do número de sensores por habitação poderá condicionar a utilização normal da internet, exigindo assim, por vezes, uma necessidade de reforço da infra-estrutura *wireless*, especialmente se falarmos de dispositivos que conectam diretamente com a internet.

A viabilidade da utilização de sensores em habitações com idosos tem vindo a ser estudada e revela-se possível. [Choi et al., 2020], após estudar a viabilidade da utilização de sensores num grupo de estudo de habitações, concluiu que, para além de ser viável, os utilizadores demonstraram ainda interesse e curiosidade em poderem ter acesso à informação recolhida pelos diferentes sensores, nomeadamente dados referentes à monitorização do seu estado de saúde.

## 3.2 *Smart Home*

*Smart home*, ou o conceito de casa inteligente, tem origem nas primeiras abordagens com sistemas de controlo de energia que poderiam ser operados remotamente.

A utilização eficiente de recursos energéticos como a eletricidade, representou claramente o mote para a proliferação destas soluções nos mais diferentes mercados [Nižetić et al., 2020]. Primariamente, estas soluções, realizavam tarefas maioritariamente induzidas ou solicitadas pelo ser humano e não por algoritmos inteligentes ou com base em dados recolhidos. Com o enorme volume de informação gerada por cada um dos sensores, podemos concretamente falar em cenários inteligentes como as *Smart Homes*, ou seja casas que, em função da informação de que dispõem, podem passar a tomar decisões de forma autónoma.

As *Smart Homes* têm vindo a ganhar especial interesse dado o baixo custo dos sensores [Gazis and Katsiri, 2021] e das possibilidades e cenários de utilização destas mesmas tecnologias. A instalação destes sensores de forma massiva e, por vezes, de livre acesso, esconde um potencial enorme no que diz respeito aos dados que estão a ser gerados [Forbes et al., 2020]. A maioria dos sensores gera dados de forma regular e quase constante que, uma vez tratados e analisados, podem servir de *triggers* para desencadear ações.

Hoje podemos encontrar múltiplos cenários de *Smart Homes* nomeadamente de poupança de energia [El-Azab, 2021], de segurança da habitação e monitorização de intrusões [Dahmen et al., 2017], sistemas integrados de análise da produção de energia através de painéis solares [El-Azab, 2021], monitorização de indicadores de saúde [Bhowmik et al., 2022] ou até sistemas de melhoria na acessibilidade e qualidade de vida [Furszyfer Del Rio et al., 2020]. A maioria das funcionalidades associadas a estes sistemas encontram-se pré programadas e o utilizador limita-se e efetuar pequenos ajustes à sua realidade e a usufruir da solução. Estima-se que, em cada

habitação, o número de dispositivos que estão conectados à internet cresça rapidamente [Zikria et al., 2021]. Com o aparecimento de tecnologias de baixo custo, nomeadamente sensores para uso doméstico, proliferam as instalações de redes de dispositivos IoT. Os sensores de IoT podem ser utilizados para diversos fins tais como: sistemas de deteção de intrusão, sistemas de domótica, sistemas controlo da habitação, entre outros [Mahmood et al., 2020].

Hoje são vários os fabricantes de soluções para *Smart Homes*, baseados em protocolos standard e, muitas vezes, até interoperáveis. O acesso e a configuração destas solução está cada vez mais facilitado, colocando as plataformas ao alcance de muitos utilizadores sem que seja necessário nenhum conhecimento técnico específico.

### 3.3 Smart Home Privacy

Quando abordado o tema da monitorização de pessoas, inevitavelmente surge a temática da privacidade na recolha e no armazenamento dos dados. No ecossistema das IoT existem variados tipos de dispositivos capazes de recolher informação, mais e menos intrusivos, tal como se verifica na figura 3.1 [Xu and Pombo, 2019].

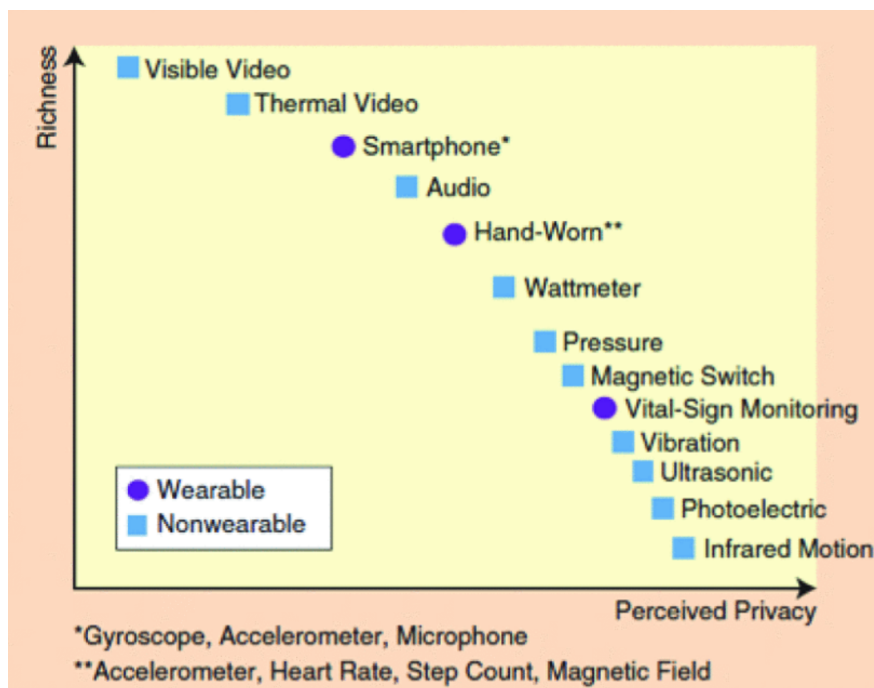


Figura 3.1: Dispositivos - Informação fornecida vs Privacidade [Debes et al., 2016]

Esta variável, influencia claramente a adoção ou não dos sistemas de monitorização. Estes sistemas suscitam frequentemente dúvidas do ponto de vista da falta de privacidade e de riscos acrescidos, nomeadamente a intrusão planeada das habitações graças à análise dos dados de ocupação. Contudo, grande parte dos utilizadores vê

uma mais valia nos sistemas de monitorização, valor este que se sobrepõe aos riscos associados à hipotética falta de privacidade [Jo et al., 2021]. Neste projeto, e de forma a mitigar as questões relacionadas com a privacidade, foram utilizados sensores *Passive Infra Red sensor*. Tal como se pode verificar na Figura 3.1, os sensores *Infrared Motion* são precisamente os que garantem maior privacidade.

### 3.4 Sensores

Os sensores utilizados para a concretização do projeto foram os sensores PIR Xiaomi®. Estes sensores de baixo custo permitem a deteção de movimento numa distância de até 7 metros (ver Figura 3.2), e num ângulo de 170° (ver Figura 3.3), Os sensores não necessitam de uma ligação à corrente elétrica de 220v, estando dotados de uma bateria que lhes confere autonomia por um período de tempo de aproximadamente um ano. Estes sensores comunicam com um agregador de informação, um *hub* (do mesmo fabricante) através da tecnologia Zigbee. O *hub* deverá estar instalado numa localização central de forma a conseguir estabelecer a comunicação com todos os sensores. Os sensores Zigbee conseguem comunicar com a unidade central até uma distância máxima de 100 metros [Kuzminykh et al., 2017].

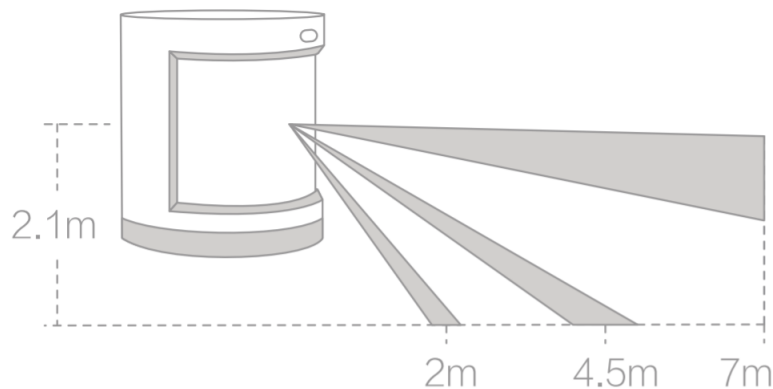


Figura 3.2: Sensor PIR [Aqara, 2021]

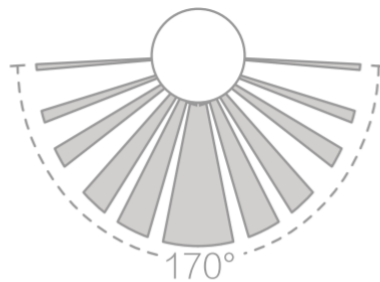


Figura 3.3: Ângulo máximo do sensor PIR [Aqara, 2021]

Estes sensores não requerem conhecimento técnico prévio para a sua instalação. O utilizador apenas necessitará de ter um telemóvel com a aplicação proprietária da marca para associar os sensores ao *hub*. Os sensores devem ser distribuídos pelos diferentes compartimentos de forma a possibilitar a recolha de informação. Cada sensor tem uma designação que deverá corresponder ao compartimento onde ficou instalado. Identificados os compartimentos com maior permanência relatada pelos habitantes, nomeadamente a cozinha, a sala de estar, o quarto e o quarto de banho, optou-se por aí instalar cada um dos sensores. Estes foram colocados de forma estratégica cobrindo com exatidão as potenciais áreas de movimento e de forma a não proporcionar o acesso fácil e conseqüente alteração do posicionamento inicial. Para o correto funcionamento da rede de sensores, o *hub* que atua como unidade central e onde todos os dados são registados deverá estar permanentemente conectado à corrente elétrica bem como à rede Wifi. Os sensores PIR, para além da sua extrema facilidade de reposicionamento, como não registam imagens dos utilizadores, tornam-se claramente não invasivos podendo a sua localização ser alterada a qualquer momento. Com o passar do tempo foi observável o esquecimento por parte dos utilizadores da presença dos sensores. Os sensores utilizados foram concebidos para utilização *indoor* e protegidos por uma estrutura em pvc resistente e de cor branca. Durante o período de duração da bateria dos sensores, um ano, não é expectável nenhuma intervenção de manutenção.

### 3.5 *Home Automation*

Ao invés das soluções proprietárias, as soluções de acesso livre, que permitem uma maior customização dos processos de automação, têm vindo a ganhar adeptos, o que, por sua vez, estimula o aparecimento de diferentes soluções integradoras que permitem aos utilizadores tirar mais partido dos dispositivos que já possuem.

*Home automation* define o conjunto de tecnologias e dispositivos necessários e utilizados para a criação de cenários de automação de rotinas na habitação. Os cenários de *Home automation* mais frequentes, estão relacionados com a otimização dos gastos de energia, nomeadamente através da gestão automática da iluminação da habitação, ou, grande parte das vezes, com a automatização de tarefas recorrentes do dia a dia, como o abrir e fechar de persianas ou mesmo ativar ou desativar eletrodomésticos remotamente. A criação de cenários de automatização representa claramente a vontade de poder controlar vários dispositivos de diferentes fabricantes de uma forma integrada para conseguir responder a situações específicas. A total customização só é possível através do desenvolvimento à medida, ou através da utilização de plataformas proprietárias que permitam o acesso aos dados. As plataformas proprietárias, nomeadamente Apple Home [Apple, 2022], Google Home [Google, 2022], Amazon Smart Home [Amazon, 2022], Shelly Cloud [Cloud, 2022],

Wink [Wink, 2022], consistem essencialmente em eco-sistemas de gestão de hardware, tais como lâmpadas, electrodomésticos entre outros. A plataforma Apple Home, baseia-se na disponibilização de uma aplicação móvel que possibilita a orquestração de vários dispositivos, certificados, através da configuração de rotinas e ações. Já a plataforma Google Home, faz uso dos seus assistentes virtuais para controlar os dispositivos da habitação. Estas plataformas, distinguem-se pela função integradora que oferecem aos seus utilizadores, ao possibilitarem a orquestração de diferentes componentes de hardware de diferentes fabricantes. Outros fabricantes disponibilizam também soluções de smart home, contudo com foco nos seus próprios componentes de hardware. Exemplo são as plataformas Shelly e Wink.

Existem ainda plataformas *Open Source* desenvolvidas para agilizar a criação de soluções customizadas. Estas plataformas contam com a participação de comunidades de *developers* que desenvolvem integrações. Exemplo dessas integrações são as *bridges* para a conexão com dispositivos, Philipps, Xiaomi, LG, entre outros fabricantes. Como exemplos de soluções de *home automation*, podemos identificar Openhab [OpenHab, 2021], homeassist [Home-Assistant, 2021], ioBroker [ioBroker, 2022], jeedom [Jeedom, 2022] ou Domoticz [Domoticz, 2022]. Duas destas principais plataformas são [OpenHab, 2021] e [Home-Assistant, 2021]. Estas soluções procuram integrar com o maior número possível de fabricantes e pretendem dar aos proprietários dos sensores e hubs uma total liberdade na sua utilização.

A *Smart Home* tem vários níveis de evolução e de automação. É certo que cada cenário terá um propósito, ainda assim é possível classificar o grau de automatização de uma smart home em seis diferentes níveis, visíveis na figura 3.4.

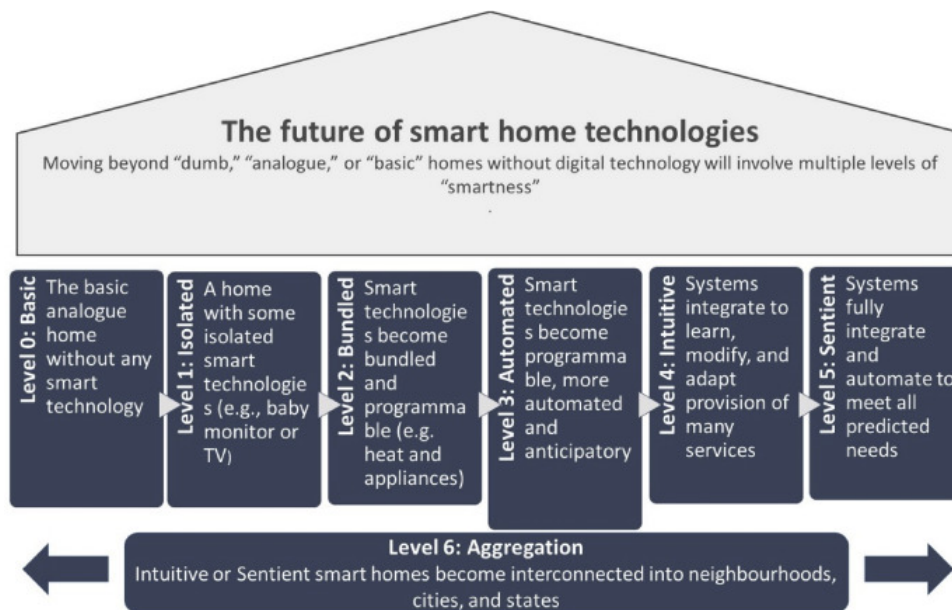


Figura 3.4: Smart Homes [Sovacool and Furszyfer Del Rio, 2020]

Os seis níveis de classificação de uma casa inteligente referidos no trabalho de [Sovacool and Furszyfer Del Rio, 2020], podem descrever-se da seguinte forma: o primeiro nível remete-se para as habitações onde se observa a introdução de tecnologia para o auxílio em tarefas, tais como a monitorização de crianças através de câmaras de vídeo ou simples lâmpadas que ligam e desligam com a presença humana; o segundo nível representa os cenários onde encontramos já dispositivos com a capacidade de serem programados ou operados à distância; o terceiro nível conta com cenários de integração de vários dispositivos e a capacidade de programar os vários para que funcionem em conjunto; o quarto nível, introduz o tema da aprendizagem, ou seja a capacidade de com os dados gerados, se possa influenciar na realização de certas actividades, exemplo em função da precisão to tempo ligar e desligar equipamentos de aquecimento; quinto nível, a *smart home* totalmente integrada será capaz de se adaptar em pleno á situação e até de prever situações e com isso controlar o funcionamento dos diferentes equipamentos; por fim o sexto e ultimo nível onde se prevê que a integração e modo de funcionamento vá para além da total automação da habitação e passe a funcionar num ambiente de integração com outras habitações, vizinhas.

## 3.6 Python

O Python [Python, 2022] é uma linguagem de programação frequentemente utilizada em processos de ETL. Dada a sua versatilidade e independência do sistema operativo, optou-se pela utilização desta linguagem nos vários processos, tanto de transformação de dados como de integração dos dados, análise estatística e até treino do modelo de ML. O Python possui um vasto portfolio de *packages*. Para o projeto em concreto, foram incluídos os seguintes *packages*:

- **pymongo** - Api Python para integração com o MongoDB;
- **json** - Manuseamento de objetos no formato JSON, nomeadamente os *documents* do MongoDB ;
- **numpy** - Operações matemáticas;
- **pandas** - Manuseamento do *dataset*;
- **csv** - Manuseamento de ficheiros CSV, nomeadamente os *datasets*;
- **matplotlib** - Representação gráfica dos valores registados;
- **datetime** - Manuseamento de *timeframes*.

### 3.7 Matplot lib

Tendo este trabalho como um dos principais objetivos identificar padrões através da análise dos dados de movimento, a existência de uma ferramenta que possibilitasse a visualização destes mesmos dados, tornou-se fundamental. Dado o já existente ambiente de desenvolvimento em linguagem Python, a opção de utilização do *package* matplotlib [Hunter, 2007] foi natural. A biblioteca Matplot foi utilizada para a criação de todos os gráficos necessários para o estudo e identificação de padrões. Esta biblioteca foi ainda utilizada na componente de identificação de *outliers*, nomeadamente através dos histogramas gerados para representar as situações de *outlier*.

## Capítulo 4

# Metodologia

Neste capítulo, fundamentam-se as opções tomadas bem como as considerações necessárias para tornar possível a realização do trabalho. Ao longo do capítulo são também descritas as barreiras identificadas bem como a forma encontrada para as contornar.

### 4.1 Nota Introdutória

A revisão bibliográfica e o estudo dos diferentes ambientes e cenários de monitorização identificados no Capítulo 2, permitiram identificar os principais componentes da solução. Para tornar possível a realização do projeto implementou-se um protótipo numa habitação própria com dois habitantes. Optou-se pela utilização de sensores de presença ao invés dos sensores *wearables* dado que se pretendia criar um cenário de monitorização o menos intrusivo possível. Para garantir uma correta recolha dos dados, foi necessário instalar e distribuir os vários sensores pelas diferentes divisões bem como garantir uma adequada ligação à internet para possibilitar a recolha dos dados registados.

A implementação requereu a instalação e configuração de um conjunto de componentes. Dada a intenção de utilizar os sensores e o ambiente de *smart home* já instalado na habitação, numa primeira fase explorou-se as diferentes formas possíveis de aceder aos dados gerados pelos sensores, uma vez que o fabricante não disponibilizava uma forma de acesso aos dados. Foi possível garantir o acesso aos dados através da instalação e configuração da plataforma OpenHab<sup>®</sup>, uma solução

de acesso livre para customização de *Smart Homes*. Uma vez garantido o acesso aos dados, iniciou-se a segunda fase que consistia na extração e transformação dos dados de forma a guardar estritamente os dados necessários para análise. Para tal foi escolhido o MongoDB [MongoDB, 2022] que guardar a informação para posterior análise. Com os dados filtrados e acessíveis, procedeu-se à análise dos mesmos para a identificação de padrões de movimento. Posteriormente realizou-se o treino de um algoritmo para a identificação de *outliers*.

Foram identificados cinco princípios éticos que devem ser respeitados para a implementação de soluções de *AAL* [Panico et al., 2020]:

- Garantir a autonomia da pessoa monitorizada, uma vez que o sistema não deve interferir no normal dia a dia do habitante;
- O benefício do habitante, uma vez que a solução deverá ter o propósito de beneficiar o idoso, nomeadamente na manutenção ou melhoria da qualidade de vida;
- Garantir que não representa uma ameaça física ou psicológica;
- Garantir que apenas espelha a realidade;
- Garantir que existe rigor na forma como os dados são tratados, não fazendo uso de eventuais diferenças éticas, culturais ou religiosas.

## 4.2 Contexto observado

O contexto observado foca-se em duas dimensões: das pessoas (sujeitos observados) e a habitação.

Os sujeitos observados no cenário experimental possuem habilitações ao nível da licenciatura e encontram-se reformados no período em análise, têm idades compreendidas entre os 68 e os 70 anos, vivem sozinhos e sem visitas regulares de familiares ou outras pessoas. Mantêm um estilo de vida com poucas deslocações, apenas as estritamente necessárias, maioritariamente para aquisição de bens alimentares.

Os dados foram recolhidos entre Fevereiro e Setembro de 2021, de forma contínua. Ambos os monitorizados consentiram e aceitaram a participação neste estudo, bem como a cedência dos dados para fins de investigação académica.

A habitação encontra-se em meio rural, estando envolvida por zona verde e com outras habitações nas imediações, no distrito de Viseu, em Portugal. Nesta habitação, foi configurado o cenário de monitorização, encontrando-se equipada com um conjunto de dispositivos do ecossistema Xiaomi<sup>®</sup>[Xiaomi, 2021]. Foram utilizados vários sensores de deteção de movimento PIR para monitorizar a presença dos habitantes nos diferentes compartimentos. Cada divisão a monitorizar estava equipada com um único sensor.

### 4.3 Acesso aos dados

Tratando-se de um ambiente real, torna-se prioritário garantir o acesso aos dados. A plataforma OpenHab, foi desenvolvida com o intuito de possibilitar aos proprietários das *smart homes*, uma verdadeira experiência de customização dos seus ambientes, bem como um acesso de forma livre aos dispositivos, ultrapassando desta forma eventuais bloqueios ou limitações de acesso. Optou-se pela utilização da plataforma OpenHab por possuir integrações nativas com os componentes do fabricante Xiaomi. Esta plataforma necessita de um sistema operativo de base para que possa ser instalada e configurada. Também com o intuito de minimizar custos e complexidade de *hardware*, a plataforma OpenHab foi instalada num micro computador Raspberry PI [RaspberryPI, 2022] (ver Figura 4.1).



Figura 4.1: Acesso aos dados dos sensores via OpenHab

A configuração da plataforma pode ser feita através do *interface web*, bem como através de API. Existem três componentes essenciais a ter em conta na utilização da plataforma, *Things*, *Channels* e *Items*, conforme apresentado na Figura 4.2. O dispositivo físico é representado no sistema por *Thing*. Este dispositivo, disponibiliza um conjunto de funcionalidades que são acedidas através dos *Channels* e, por fim, as funcionalidades, que são representadas pelos *Items*.

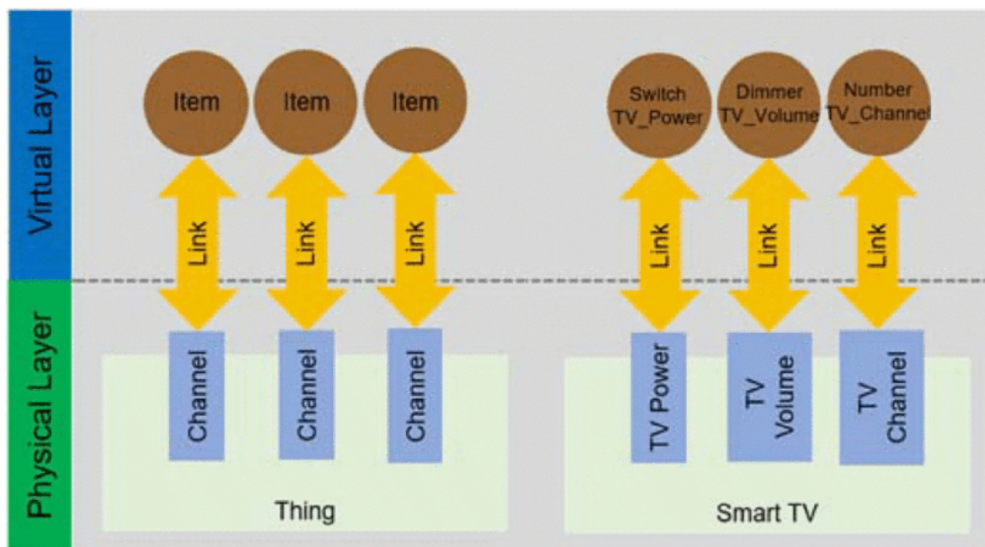


Figura 4.2: OpenHab *components* [Parocha and Macabebe, 2019]

#### 4.4 Extração e pré-processamento de dados

O processo de extração dos dados foi simplificado pela utilização da solução OpenHabian (OpenHab e Debian) uma vez que já existem conectores nativos para integrar com os dispositivos IoT, nomeadamente o *Hub* Xiaomi. Todos os dados gerados devem ser mantidos e nunca apagados de forma a tornar o dataset tão rico quanto possível. A estratégia de *Extract Transform Load* (ETL) foi pensada de forma a evitar a perda de dados. Foi particularmente importante garantir a consistência dos dados não só na recolha como também no armazenamento.

Cada um dos sensores de movimento quando é ativado pela existência de movimento na divisão monitorizada, comunica ao *hub* essa mesma alteração de estado e essa informação fica registada em formato de *log*, com o nome do sensor, a data e hora com os milissegundos em que aconteceu, sendo posteriormente necessário recolher e agrupar essa informação. No momento em que o *log* é capturado, apenas fica registado o nome do sensor e o registo temporal, evitando o registo de informação desnecessária. Todo o processamento de extração e limpeza dos dados acontece na unidade de processamento Raspberry PI, localmente, antes de se proceder ao armazenamento em MongoDB (ver Figura 4.3).

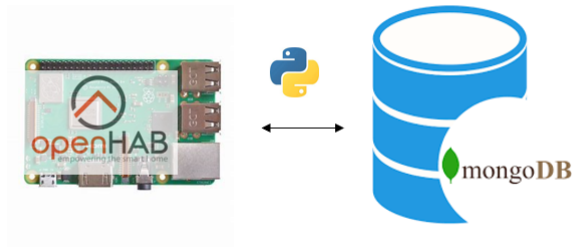


Figura 4.3: Carregamento dos dados para MongoDB

O fluxo de extração, transformação e carregamento dos dados acontece em intervalos de 60 segundos. O curto período de tempo foi uma opção tendo em conta a deteção de algumas falhas de comunicação e energia, tentando mitigar desta forma a falha no registo da informação.

## 4.5 Dataset

Dada a necessidade de armazenar um número considerável de dados de forma recorrente e sem uma noção exata do espaço necessário, foi opção para este projeto a utilização de uma base de dados de fácil acesso e sem necessidade da criação de um modelo relacional de dados. Assim, optou-se por um motor de dados *Non Relational Database* (non-SQL) do fabricante MongoDB [MongoDB, 2022]. A utilização de uma base de dados non-SQL para armazenamento dos dados permite a flexibilidade da base de dados, no que diz respeito à capacidade de registar dados bem como retornar dados quando efetuadas consultas. Dada a flexibilidade da linguagem Python e a sua utilização no manuseamento dos dados, optou-se pela utilização da API em Python (pymongo). Com esta API, foi possível guardar dados e fazer as consultas necessárias para garantir a consistência, nomeadamente, em caso de falha, garantir a não duplicação dos mesmos.

Qualquer modelo de *Machine Learning*, necessita de um bom *dataset*. Inevitavelmente, os resultados finais, nomeadamente a deteção de *outliers*, serão influenciados pela qualidade do *dataset*. Para este projeto, os dados tiveram de ser estudados e tratados para permitir uma melhor interpretação dos mesmos.

Por se tratar de dados registados ao longo do tempo e com importância do ponto de vista temporal, foi necessário identificar e estudar ferramentas que possibilitassem agrupar os dados, por diferentes períodos de tempo ( hora, dia, semana, mês, consoante o estudo a realizar). Os dados foram trabalhados de forma a poderem ser utilizados, isto é, foram convertidos num *dataframe*. Dado que são gerados muitos dados em pouco tempo, a análise de registo a registo não se tornou relevante, tendo

sido por isso necessário avaliar estratégias para lidar com séries de dados temporais. O *dataset* original serviu ainda como ponto de partida para a criação de um *dataset* rico em *features* de modo a possibilitar o treino do algoritmo para detecção de *outliers*.

Relembrando que o projeto assenta num estudo e caracterização de dados recolhidos de uma rede de sensores instalada numa habitação de idosos ao longo de vários meses e 24 horas por dia, rapidamente se percebe que estamos a falar de um conjunto de dados observados de forma contínua ao longo do tempo. Regra geral o *data frame*, estrutura em python que contém vários registos indexados pela data e hora do acontecimento, é utilizado precisamente para o estudo de padrões. Os *datasets* em formato de *time series* têm uma vasta aplicabilidade em cenários do mundo real, nomeadamente no campo da geologia na predição de sismos e erupções, na área financeira, na previsão do tempo, entre outros [Faouzi and Janati, 2020].

## 4.6 Detecção de Anomalias

A detecção de anomalias em conjuntos de dados, tem sido um tópico bastante ativo no estado da arte, capítulo 2, dada a sua relevância em diferentes contextos, desde a economia [Domashova and Kripak, 2021] à biologia [Selicato et al., 2021], saúde [Rajeswari et al., 2019] etc. Tradicionalmente em ambientes fabris com o intuito de otimizar operações, detetar antecipadamente eventuais falhas e maus funcionamentos, ou ainda na área da banca, na detecção de fraude, implementam-se estratégias para identificação de cenários extremos ou, por outras palavras, detecção de *outliers*.

Existem várias técnicas para a detecção de *Outliers* [Wang et al., 2019], nomeadamente:

- A utilização de dados estatísticos por norma baseados em regressões lineares e consequente criação de intervalos de valores: se o valor registado não estiver no intervalo apresentado, pode então tratar-se de um *outlier*;
- *Clustering* de pontos, agrupando os pontos registados por proximidade, criando desta forma zonas ditas normais e, assim, os valores não pertencentes aos *clusters* são vistos como *outliers*;
- Classificação das ocorrências e consequente treino de situações normais e anómalas;
- O ponto vizinho mais próximo, calculando a média da distância dos pontos, aferindo desta forma um valor de distancia ou proximidade.

Neste domínio, a inteligência artificial e os modelos desenvolvidos, assumem um papel importante por tornarem possível o treino de modelos de ML que, com os dados gerados pelos sensores, podem identificar situações de risco em tempo útil.

De forma a concretizar a deteção de anomalias, foi necessário estudar e avaliar a *performance* e verdadeira capacidade do algoritmo de *Machine Learning, isolation forest* [Liu et al., 2008]. Ainda que os modelos sejam treinados para melhor executarem uma sinalização dos *outliers*, o contexto mereceu uma análise estatística inicial ao invés de numa implementação direta de um modelo de ML. Os algoritmos utilizados para realizar este tipo de deteção e alterações aos normais padrões, podem ser supervisionados e não supervisionados [Wang et al., 2019]. Para a realização do projeto optou-se pelo treino de um algoritmo não supervisionado, *isolation forest* [Liu et al., 2008].



## Capítulo 5

# Implementação

Este capítulo apresenta a arquitectura e a implementação do projeto. Na arquitectura são apresentados os diferentes componentes utilizados no projeto. Na implementação é feita a contextualização das opções tomadas no que diz respeito aos sensores e á sua localização, a descrição dos mecanismos de ETL, os mecanismos adotados para o estudo de padrões e por fim a abordagem para o treino do modelo de *Machine Learning*.

### 5.1 Arquitectura

A arquitectura adotada para a aquisição de dados é composta pelos sensores de movimento, pelo *hub*, que recebe a informação dos sensores, e é acedido via Wifi pelo dispositivo Raspberry PI, em seguida os dados são armazenados num ambiente MongoDB tal como ilustrado na Figura 5.1.

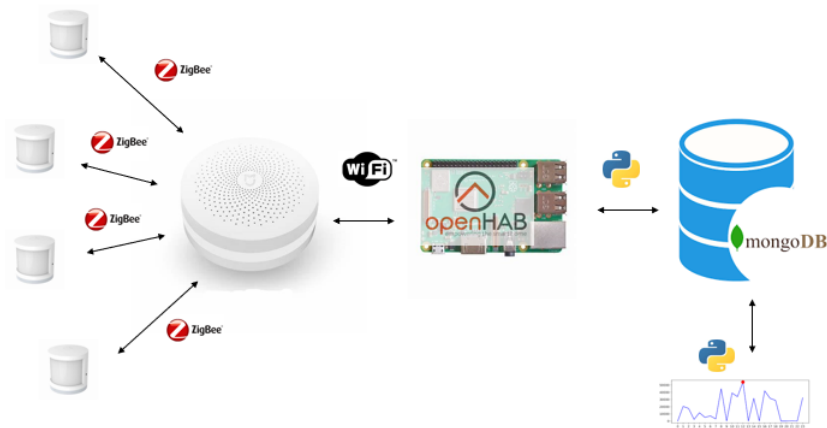


Figura 5.1: Arquitetura da solução de aquisição e integração de dados

Para o projeto foram utilizados quatro sensores de movimento e um *hub*. A conexão entre os dispositivos e o *hub* é pré-estabelecida através da tecnologia Zigbee. Os sensores foram instalados em quatro divisões diferentes, de forma a poderem monitorizar as movimentações dos habitantes em cada um dos compartimentos, num regime contínuo.

O *hub* não expõe diretamente os dados que recolhe, tendo sido necessário avaliar soluções de *home automation*. Optou-se por uma solução *open source* responsável por garantir o acesso aos dados. O OpenHab foi a plataforma escolhida pelo desenvolvimento e suporte a conectores nativos com os *hubs*. Para o projeto, foi ainda adquirido um *mini PC* (*Raspberry PI*), onde foi feita a instalação de uma imagem com o OpenHabian, uma implementação linux do OpenHab.

A arquitetura pressupõe a existência de uma ligação à Internet para possibilitar o armazenamento dos dados num repositório MongoDB.

Neste projeto existem duas redes de comunicação: Zigbee – para ligação entre sensores e *hub* – e Wifi – para ligação entre *hub* e Raspberry PI. A integração com MongoDB foi estabelecida através da integração desenvolvida na linguagem Python [Python, 2022].

## 5.2 Localização dos sensores

Os sensores utilizados para este projeto piloto têm uma dimensão reduzida ( $30\text{mm} \times 30\text{mm} \times 33\text{mm}$ ), e por essa mesma razão facilmente, depois de estarem instalados, passam despercebidos nos compartimentos monitorizados.

A colocação dos sensores nos diferentes compartimentos foi pensada de forma a maximizar a cobertura do espaço a monitorizar. Tendo em conta que os sensores são capazes de captar movimento num ângulo de  $170^\circ$ , a colocação dos sensores deverá

ter em conta as principais zonas de movimento dentro do compartimento onde se encontram.

Foram escolhidos os compartimentos onde as pessoas realizam as suas actividades principais no seu dia a dia. Evitaram-se assim zonas de transição entre compartimentos, assim como zonas que não estejam associadas a actividades importantes para os objectivos do estudo. Foram escolhidos quatro compartimentos e foi colocado um sensor em cada um deles: quarto (Figura 5.2), casa de banho (Figura 5.3), sala (Figura 5.4), e cozinha (Figura 5.5).

Pelo facto de serem sensores alimentados apenas a bateria, é fácil posicioná-los em qualquer localização do compartimento, uma vez que não estão condicionados à existência de corrente eléctrica nas proximidades. Foram considerados apenas locais de instalação de difícil acesso, de forma a evitar a deslocação do sensor e, por sua vez, falha de registo de dados.

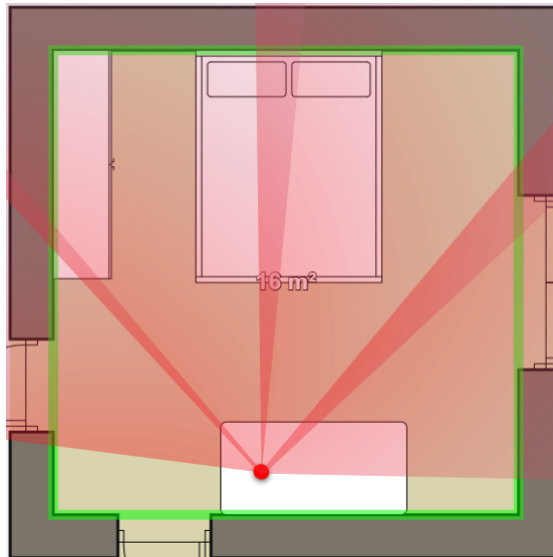


Figura 5.2: Quarto (Sensor 1)

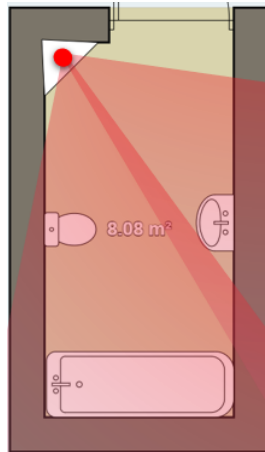


Figura 5.3: Casa de banho (Sensor 2)

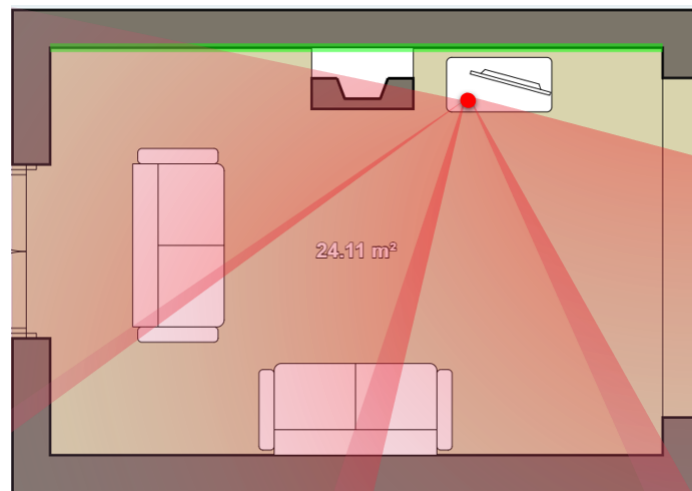


Figura 5.4: Sala de estar (Sensor 3)

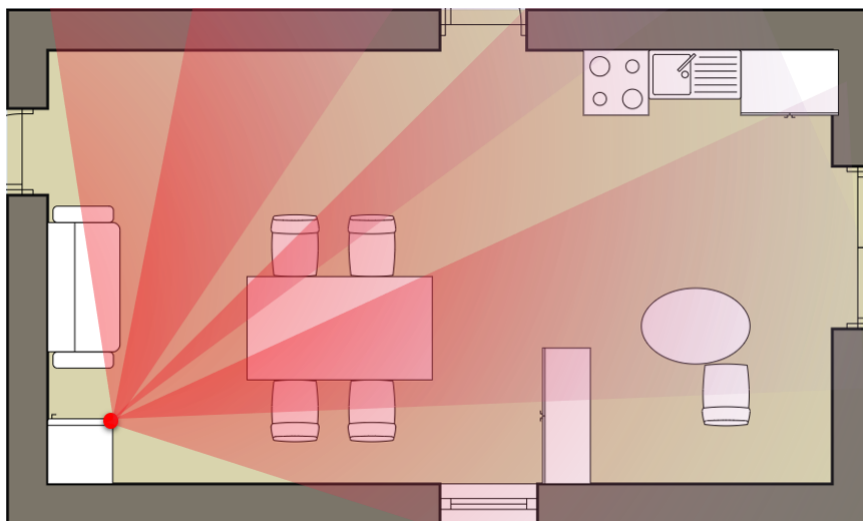


Figura 5.5: Cozinha (Sensor 4)

## 5.3 Recolha de dados

Após a instalação dos diferentes sensores e garantindo a sua conexão com o *hub*, foi necessário encontrar uma forma de obter os dados ou registos dos sensores. Atendendo a que o *hub* é de um fabricante que não expõe uma API ou outro tipo de acesso para a extracção dos dados, optou-se pela configuração e instalação de uma solução de *middleware* capaz de aceder ao *hub* e aos seus dados. A solução OpenHab [OpenHab, 2021], disponibiliza um conjunto de conectores para diferentes tipos de dispositivos IoT (ver secção 3.5), funcionando desta forma como *middleware* entre o *hub* e a plataforma de dados online.

Tal como descrito na secção 3.2, para a instalação da plataforma OpenHab foi utilizada uma imagem Linux OpenHabian. De forma a minimizar a necessidade de configurações na instalação, o Raspberry PI foi pré-configurado para utilizar a mesma rede Wifi onde o *hub* já estava configurado.

Após a instalação desta solução, foi feito um varrimento para encontrar dispositivos passíveis de serem integrados, nomeadamente os sensores PIR. Foi configurada a *binding* que permitiu estabelecer o contacto com o *hub* e, por cada *Thing* (sensor), um *Channel* para *logging* das ocorrências geradas pelo sensor, conforme descrito em 4.3. Os logs ficam armazenados no formato de ficheiro de texto e escritos de forma incremental facilitando a sua leitura.

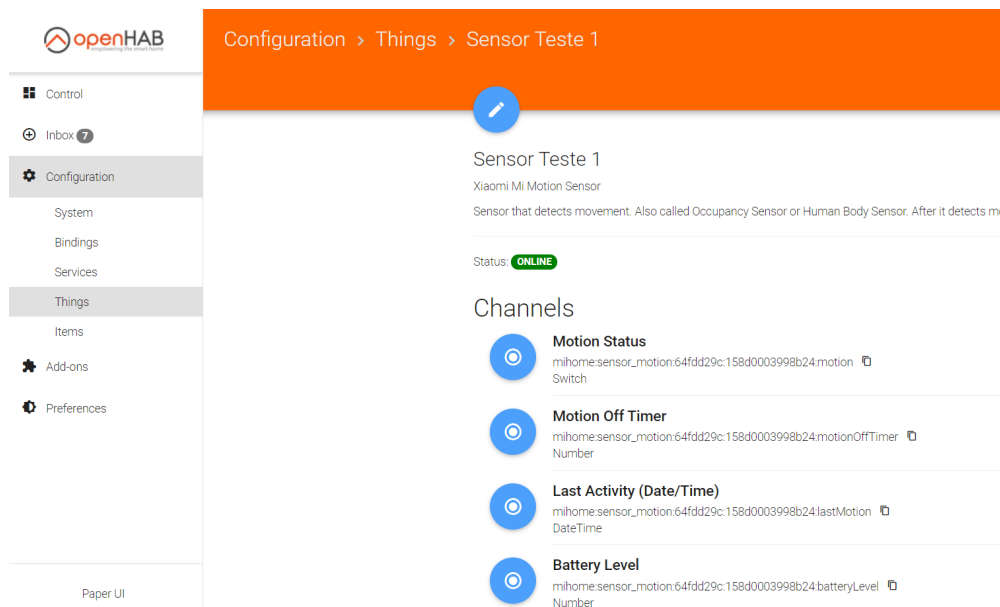


Figura 5.6: Portal OpenHab®

## 5.4 Integração dos dados

A plataforma OpenHab disponibiliza uma imagem Debian Linux com as configurações iniciais necessárias para a utilização da plataforma. Esta imagem foi instalada

num dispositivo Raspberry PI versão 3 com 512 mb de RAM.

O *hub* e o Raspberry encontram-se ambos conectados à mesma rede Wifi. O software necessário para a execução de scripts python [Python, 2022] bem como as respetivas dependências foram configurados nesta unidade. Os scripts python que são executados nesta unidade são responsável por realizar, desde a extração dos dados, leitura do ficheiro de *log*, à sua transformação, limpeza das *strings* de forma a importar apenas o estritamente necessário. Esta unidade é ainda responsável pela execução de *scripts* para salvar os dados, nomeadamente através da utilização da API do MongoDB para a criação de novos documentos.

As rotinas de sincronização dos novos *logs* gerados para a plataforma MongoDB foram desenvolvidas de forma a garantir a não duplicação dos dados. Os *logs* são limpos e adicionados ao MongoDB de forma incremental através da leitura e comparação com o último registo adicionado. Foram feitas configurações para que, em caso de falha e restabelecimento de energia, todos os processos voltem ao seu funcionamento normal.

Os *logs* capturados pelo OpenHab contêm muita informação e podem representar diferentes eventos, de vários tipos. Podemos encontrar nos *logs* eventos associados à deteção de movimento (ver Figura 5.1), bem como as ações de ativação e desativação do próprio sensor. Foi necessário, numa primeira abordagem, validar quais os *logs* relevantes para o projeto. O *script* Python responsável pela integração dos dados é executado com a periodicidade de um minuto para gravar no MongoDB os valores já organizados (ver Figura 5.2). A API é explorada com duas finalidades a procura de registos e a inserção de registos.

Tendo em conta que os registos dos movimentos de sensores possuem sempre uma data e uma hora, garante-se desta forma a unicidade dos registos, especialmente porque o valor da hora contém as horas, os minutos, segundos e milésimos de segundo.

Os dados são carregados para o MongoDB de forma sequencial e não têm qualquer tipo de ordenação, existem num único espaço. A estrutura de dados no motor MongoDB respeita um formato *Java Script Object Notation* (JSON).

---

```
1      2021-01-31 18:24:29.312 [vent.
      ItemStateChangedEvent]
      XiaomiMiMotionSensor1_
      LastActivityDateTime changed from
      2021-01-31T18:23:11.783+0000 to
      2021-01-31T18:24:29.308+0000
```

---

*Código fonte 5.1: OpenHab log*

A Figura 5.2 apresenta um exemplo de um registo guardado na base de dados MongoDB. O formato do registo contém os campos: *oid* (*object identification*), a data, a hora, o ID do sensor e o evento ocorrido.

```
1      "$oid": "60284c135ed15f380a0c35ab",  
2      "date": "2021-01-30",  
3      "time": "17:07:55.544",  
4      "sensorID": "XiaomiMiMotionSensor4",  
5      "event": "ON"
```

Código fonte 5.2: Registo do evento gerado no processo de integração

## 5.5 Caracterização do dataset

Os dados que compõem o *dataset* foram recolhidos durante um período de oito meses (Fevereiro a Setembro de 2021), de forma diária. Existem ausências de registos justificadas por quebras de energia elétrica e pela necessidade pontual de reiniciar os equipamentos.

O *dataset* global original contém 60689 registos com as seguintes colunas: *data hora*, *compartimento* e *disparo*, tal como apresentado na Figura 5.7. Cada registo representa uma ativação de um determinado sensor num determinado momento.

O *dataset* é composto por dados registados cada vez que os sensores são ativados sem qualquer tipo de agrupamento. Este *dataset* foi criado incrementalmente, adicionando registos à base de dados MongoDB tendo sido posteriormente exportados para um ambiente local onde foi feito o seu pré-processamento.

```
2021-02-20 22:44:33,WC,1  
2021-02-20 22:45:22,LivingRoom,1  
2021-02-20 22:45:34,WC,1  
2021-02-20 22:48:01,Kitchen,1  
2021-02-20 22:49:40,Kitchen,1  
2021-02-20 22:51:22,Kitchen,1  
2021-02-20 22:52:22,Kitchen,1  
2021-02-20 22:54:29,Kitchen,1  
2021-02-20 22:55:52,Kitchen,1  
2021-02-20 22:59:49,LivingRoom,1  
2021-02-20 22:59:50,WC,1  
2021-02-20 23:01:28,LivingRoom,1  
2021-02-20 23:02:56,LivingRoom,1  
2021-02-20 23:04:18,LivingRoom,1
```

Figura 5.7: Excerto do dataset original

## 5.6 Pré-processamento de dados

Para proceder à identificação de padrões nos dados recolhidos, foi necessário organizar os dados.

O pré-processamento dos dados consistiu em:

- Segmentar o *dataset* de forma a criar um ficheiro por compartimento;
- Remover dias com registos em falta;
- Agrupar temporalmente os registos.

Os dados gerados pelos sensores encontram-se ordenados cronologicamente. Para realizar a análise individual dos padrões de movimento em cada um dos compartimentos, foi necessário proceder à organização dos dados, separando os registos referentes a cada um. A divisão dos dados permitiu iniciar o estudo dos padrões compartimento a compartimento sem a complexidade dos restantes registos.

### 5.6.1 Eliminação de períodos incompletos

O estudo dos *datasets* individuais permitiu identificar a existência de falhas no registo de dados. Foi também confirmado junto dos habitantes a incoerência dos registos, confrontando a falta de resultados com a confirmação de presença na habitação.

Apurou-se um valor mínimo de registos representativo de um dia completo como referência, através da comparação dos registos totais diários. Desta forma, foram apenas considerados os dias com um número total de ocorrências superior ao valor de referência para cada compartimento. Para a cozinha considerou-se um mínimo de 200 movimentos por dia; para a sala 70, e 20 para a casa de banho e quarto. Estes valores mínimos obtiveram-se pela observação dos dias em que foi possível identificar registos aos longo das 24 horas do dia. Assim, os dias com registos contínuos foram considerados dias de referência no que diz respeito ao número de registos.

### 5.6.2 Agrupamento de dados

O agrupamento temporal de dados tem um impacto na detecção de padrões comportamentais. Se a granularidade temporal for muito reduzida, os *outliers* podem comprometer a identificação de regularidades nos dados. Quando demasiado elevada, muitos dos padrões são absorvidos pela agregação dos dados [Rawassizadeh et al., 2019].

Os valores recolhidos diariamente foram agrupados em períodos que fizessem sentido. O intervalo de tempo de uma hora foi o período inicialmente escolhido. Torna-se mais visível o padrão quando os dados estão precisamente agrupados em períodos de uma hora. Outros períodos foram também explorados, contudo períodos mais curtos apresentavam muito ruído, períodos maiores representavam dados mais

abstratos e com pouca utilidade, nomeadamente para o estudo que se pretendeu realizar. Os dados foram segmentados pelas diferentes horas do dia, garantindo assim que era feita a avaliação do padrão num período útil, dado que, avaliando as 24 horas do dia, conseguimos avaliar um ciclo completo do dia dos participantes.

Para obter o nível apropriado de granularidade, foram exploradas ainda diferentes agregações dos dados, contudo os períodos de 30 e 60 minutos demonstraram ser os períodos de tempo que melhor permitem a visualização dos padrões (Figuras 6.1, 6.4, 6.2 e 6.3). O mecanismo de *resampling* tenta agrupar os dados em iguais períodos de tempo com base na configuração pretendida. No caso de não existirem registos num determinado período de tempo, o intervalo é criado com valor 0 na coluna *sensor* (número de disparos). O agrupamento do número de disparos é interpretado como o total de disparos do sensor durante o período de tempo correspondente.

### 5.6.3 Formatação de *features*

Em ML as *features* de um *dataset* podem ser quantitativas, qualitativas, discriminatórias. São representadas habitualmente por colunas no *dataset*. Na maioria das apresentações estes dados estão convertidos em algarismos.

Os dados utilizados neste estudo de identificação de padrões encontram-se no formato de séries temporais, organizados por diferentes ficheiros, de forma a facilitar o estudo dos padrões, possibilitando observar dados de forma separada.

O *dataset* utilizado para a criação dos modelos de ML obriga a várias transformações realizadas na estrutura de base dos dados, de forma a criar as *features* com o formato adequado para a criação dos modelos.

A *data*, em particular, foi decomposta em várias *features*, tais como, o dia da semana (valor numérico de 1 a 7) e a hora ( $h$ ), na representação decimal (Equação 5.1).

$$feature_h = h + minutos/60 \quad (5.1)$$

O formato do registo contempla também a transformação da ativação de presença do sensor no compartimento pela representação binária (0 e 1 representam a não ativação e ativação do sensor, respetivamente).

A Figura 5.8 mostra a representação final dos registos transformados nas *features* utilizadas para criação dos modelos de *Machine Learning*.

```

2021-02-21,02:00:00,Sun,Feb,2.0,1,0,0,0
2021-02-21,02:20:00,Sun,Feb,2.33,1,0,0,0
2021-02-21,02:40:00,Sun,Feb,2.67,1,1,1,1
2021-02-21,03:00:00,Sun,Feb,3.0,0,0,1,1
2021-02-21,03:20:00,Sun,Feb,3.33,0,0,0,0
2021-02-21,03:40:00,Sun,Feb,3.67,0,0,0,0
2021-02-21,04:00:00,Sun,Feb,4.0,0,0,0,0
2021-02-21,04:20:00,Sun,Feb,4.33,0,0,0,0
2021-02-21,04:40:00,Sun,Feb,4.67,0,0,0,0
2021-02-21,05:00:00,Sun,Feb,5.0,0,0,0,0
2021-02-21,05:20:00,Sun,Feb,5.33,0,0,0,0
2021-02-21,05:40:00,Sun,Feb,5.67,0,0,0,0
2021-02-21,06:00:00,Sun,Feb,6.0,0,0,0,0
2021-02-21,06:20:00,Sun,Feb,6.33,0,0,0,0
2021-02-21,06:40:00,Sun,Feb,6.67,0,0,0,0
2021-02-21,07:00:00,Sun,Feb,7.0,0,0,0,0

```

Figura 5.8: Excerto do *Dataset*

## 5.7 Reconhecimento de padrões

A identificação de padrões foi realizada através de três análises visuais distintas: (1) distribuição de todos os dados registados; (2) resultado do cálculo dos percentis 5 e 95; e (3) cálculo da probabilidade de ocupação.

Os padrões consistem na observação de um conjunto de características que, regra geral, tendem a repetir-se. Os dados foram carregados para uma estrutura de dados temporal (*time series*), sendo posteriormente analisados a partir da sua representação gráfica (Secção 3.7).

Para a análise visual dos padrões, foram utilizados três tipos de gráficos: (1) somatório dos disparos; (2) percentis associados aos disparos; e (3) probabilidade de ocupação temporal do espaço.

A primeira análise visual consistiu na interpretação do gráfico gerado com o somatório dos disparos para cada hora do dia, por compartimento. Desta forma foi obtida uma representação gráfica com todos valores, máximos e mínimos, por hora.

Posteriormente, foram analisados os percentis 5 e 95 (Secção 6.1), filtrando potenciais *outliers* que podem comprometer a análise de padrões. Na identificação dos percentis 5 e 95, a ausência de registos tem um grande peso no cálculo do percentil 5 dado que, sempre que não existe movimento, o mecanismo de *resampling* vai atribuir o valor de zero ocorrências para aquele período de tempo. Uma vez que o mecanismo de *resampling* recria todos os períodos temporais, seja um dia, uma hora ou trinta minutos, será necessário ter sempre em conta a criação destes períodos para dias anteriormente identificados como dias maus. Nesta situação, opta-se por transformar estes agrupamentos, que resultaram no valor 0, em NaN. Quando os valores se encontram no formato NaN, estas ocorrências não são consideradas para o cálculo dos percentis – i.e., ao traçarmos o gráfico com as ocorrências, os valores NaN não são representados.

Por fim, foram calculadas as probabilidades de disparo dos sensores de movimento por hora e compartimento. Este cálculo foi realizado através da Equação 5.2. A probabilidade é igual á contagem das horas onde o número de disparos é superior ao valor fixo (10 neste caso) dividido pelo número total de dias considerados. O valor 10 foi determinado pela análise gráfica dos padrões. Um valor inferior a 10 disparos por hora não representa que tenha existido presença no compartimento.

$$Probabilidade_h = \frac{\#(disparos_h > valor\ fixo)}{\#dias} \quad (5.2)$$

Nas três representações descritas anteriormente, o tratamento dos dados, a sua organização e transformações, foram feitos com recurso à linguagem Python 3.6.

## 5.8 Algoritmo de *Machine Learning*

Neste contexto, por não sabermos à partida identificar o que são registos *outliers* e os que não são, não é viável a utilização de um algoritmo de *Machine Learning* supervisionado.

A deteção de anomalias obriga normalmente à utilização de algoritmos de aprendizagem não supervisionada. Para a criação de modelos de *machine learning* neste trabalho, foi adotado o algoritmo *Isolation Forest*. Este algoritmo foi identificado em trabalhos semelhantes, tais como [Liu et al., 2008] e [Kurnianingsih et al., 2019].

O *isolation forest* calcula um *score* para cada registo, que representa a profundidade a que o registo se encontra na da árvore de decisão gerada pelo algoritmo. Uma vez que se trata de um algoritmo de *clustering*, ao agrupar eventos com um *score* semelhante, o algoritmo consegue identificar os *outliers* – i.e., eventos afastados dos *clusters* identificados durante a fase de treino.

A biblioteca *sklearn ensemble* foi necessária para possibilitar a implementação e treino do algoritmo. Na implementação do algoritmo *isolation forest*, foram tidos em conta os seguintes parâmetros: *n\_estimators*, onde utilizamos o valor 50 (valor de referência); no *contamination*, que representa a percentagem esperada de *outliers* no *dataset*, utilizou-se o valor de 0.10. O *dataset* utilizado e partilhado na Figura 5.8 foi dividido em dois blocos, treino e teste, tendo sido aplicada a seguinte divisão: 80% do *dataset* para treino e 20% para teste. Para dar início ao treino utilizou-se o método *fit* com o subset de dados de treino e para a validação do modelo utilizou-se o método *predict* com o subset de dados de teste.



## Capítulo 6

# Análise de Resultados

Neste capítulo é apresentado o processo de análise de dados que inclui a caracterização do *dataset*, o pré-processamento dos dados, os resultados do estudo para a identificação dos padrões, os resultados do modelo treinado para a identificação de *outliers*, e, por fim, a discussão dos resultados.

### 6.1 Identificação de padrões

A validação da existência de padrões temporais de movimentação dos sujeitos do estudo em compartimentos específicos da casa é um dos objectivos principais deste trabalho.

#### 6.1.1 Acumulado de registos

A observação dos dados acumulados permite visualizar a distribuição dos dados sem excluir nenhum tipo de registo, agrupando-os apenas por períodos de tempo, mais concretamente por períodos de 60 minutos. Todos os dados registados foram representados no gráfico do respectivo compartimento.

A *cozinha* (Figura 6.1) é o compartimento mais frequentado durante o dia, dada a quantidade de disparos identificados. Neste compartimento, os períodos de maior presença – onde se identifica um maior número de disparos do sensor – estão compreendidos entre as 7 e as 15 horas e entre as 17 e as 21 horas (Figura 6.1). A *cozinha* é o compartimento onde os habitantes permanecem durante o dia.

É observável uma maior presença na *cozinha* durante os períodos das refeições. No período compreendido entre as 14 e as 16, encontramos uma diminuição da presença que se justifica com o hábito, identificado, da sesta após almoço. Esta alteração pode ser igualmente verificada ao analisarmos os dados do compartimento *quarto* (Figura 6.4).

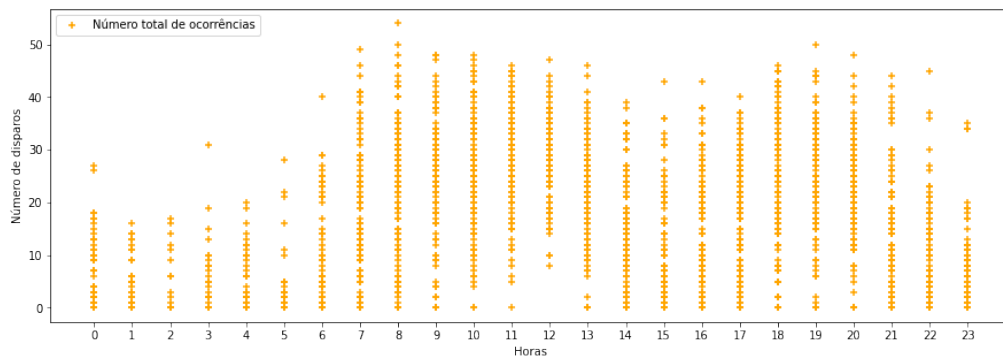


Figura 6.1: Total de ocorrências por hora no compartimento *Cozinha*

No compartimento *sala* é visível a presença no período entre as 21 e as 24 horas (Figura 6.2), enquanto se identifica uma menor presença na *cozinha* no mesmo período. Foi confirmada pelos habitantes a rotina de se deslocarem da *cozinha* para a *sala* após o jantar e a sua permanência nesse espaço até se deslocarem para o *quarto* para dormir. A *sala* é um espaço aberto que se encontra no mesmo piso dos compartimentos *quarto* e *casa de banho*, o que justifica a existências de algumas movimentações durante o período da manhã entre as 7 e as 10 horas.

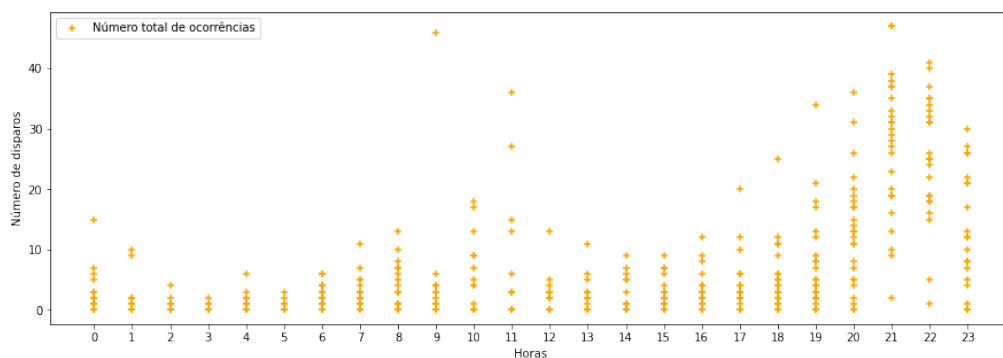


Figura 6.2: Total de ocorrências por hora no compartimento *Sala*

A *casa de banho* é um compartimento de presença tradicionalmente curta. Ainda assim é possível identificar alguns períodos recorrentes de utilização, nomeadamente o início da manhã – provavelmente relacionado com o acordar e a higiene matinal, entre as 7 e 8 horas da manhã (Figura 6.3) e das 22 e 23 horas (Figura 6.3), antes de se dirigirem para o *quarto*.

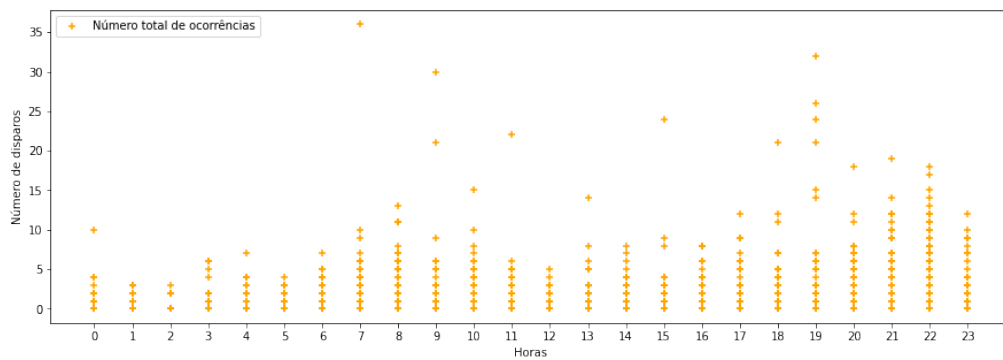


Figura 6.3: Total de ocorrências por hora no compartimento *Casa de banho*

O último compartimento analisado foi o *quarto*. Este espaço está habitualmente reservado para o repouso. É possível identificar a ocupação nos períodos da noite e da sesta (Figura 6.4). Entre as 0 e as 8 horas existe uma maior movimentação, assim como no período da sesta (entre as 14 e 17 horas) e no fim do dia (entre as 22 e 23 horas).

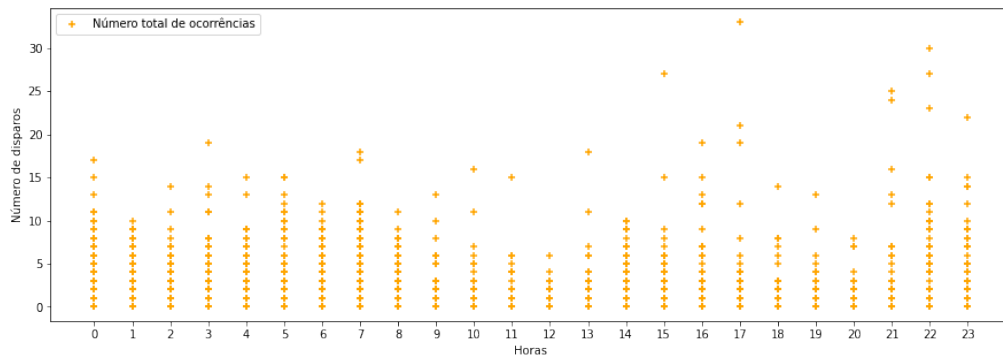


Figura 6.4: Total de ocorrências por hora no compartimento *Quarto*

### 6.1.2 Distribuição Temporal das Presenças

A análise dos totais de presenças não permite avaliar a sua distribuição, por hora, durante o período analisado. Por sua vez, os *outliers* no *dataset* de treino, podem comprometer a identificação de padrões através de uma análise do número máximo e mínimo de presenças. Uma análise de percentis pode eliminar o ruído nos dados provocado por *outliers* gerados por eventos não pertencentes à rotina das pessoas.

Depois de avaliar a distribuição dos dados através da visualização de todos o registos diários, foi feita a análise dos percentis 5 e 95. O cálculo dos percentis permitiu confirmar as constatações obtidas através da análise do acumulado de registos (6.1.1), no que diz respeito à identificação de padrões. Nesta análise, no compartimento *cozinha* (Figura 6.5), podemos identificar períodos onde existe movimento, nomeadamente entre as 10 e 13 e às 20 horas.

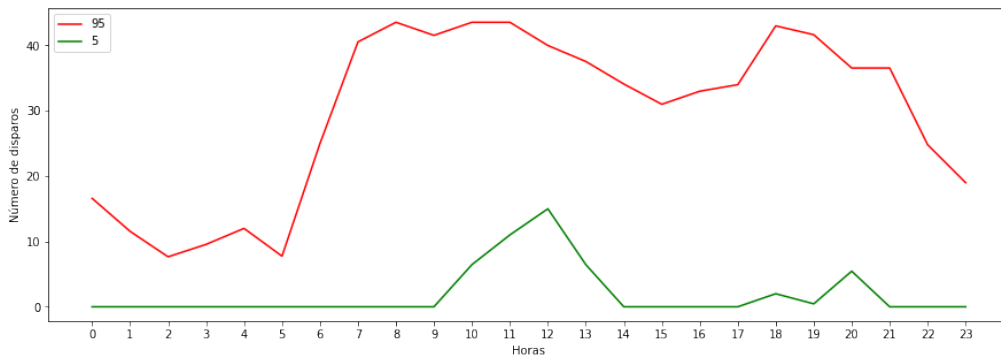


Figura 6.5: Disparos no compartimento *Cozinha*. Percentil 95 e 5

No compartimento *sala*, voltamos a validar a presença das pessoas no período entre as 21 e 22 horas (Figura 6.6). Neste período de tempo podemos observar que o número de disparos chega a ser próximo do número de disparos do compartimento *cozinha*, nos períodos mais movimentados (Figura 6.5), com valores a rondar os 40 disparos por hora.

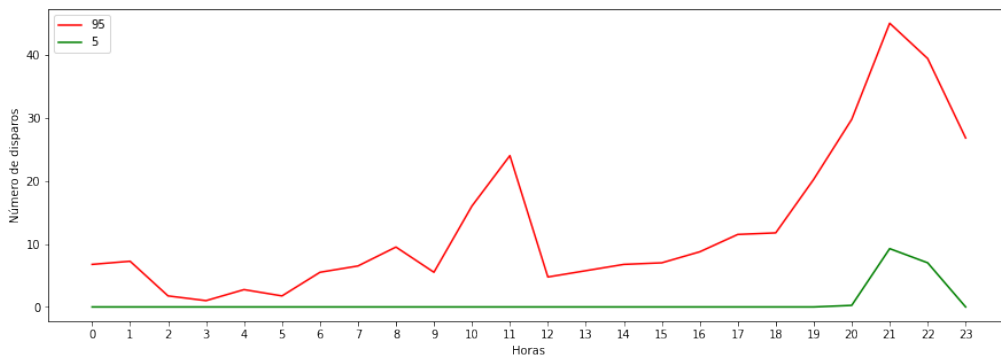


Figura 6.6: Disparos no compartimento *Sala*. Percentil 95 e 5

Na *casa de banho*, tal como expectável, não se identificam movimentações durante grande parte do dia (Figura 6.7). O percentil 95 revela uma maior presença no início da manhã (entre as 6 e as 10 horas) e no fim do dia (entre as 19 e as 23 horas), tal como ilustrado na Figura 6.7).

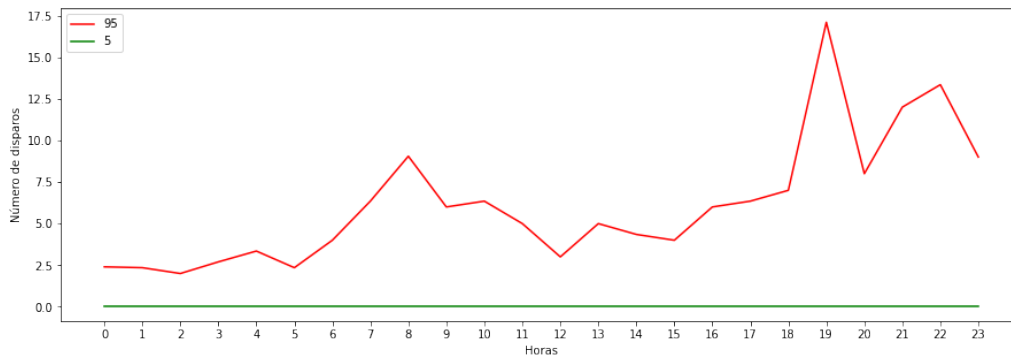


Figura 6.7: Disparos no compartimento *Casa de Banho*. Percentil 95 e 5

À semelhança do compartimento *casa de banho*, o compartimento *quarto* revela novamente a presença de padrão, nomeadamente quando analisado o percentil 95. Existe uma maior presença durante o período da noite (entre as 0 e as 6), durante o período da sesta (entre as 14 e as 16) e no final do dia (entre as 22 e as 23) (Figura 6.8).

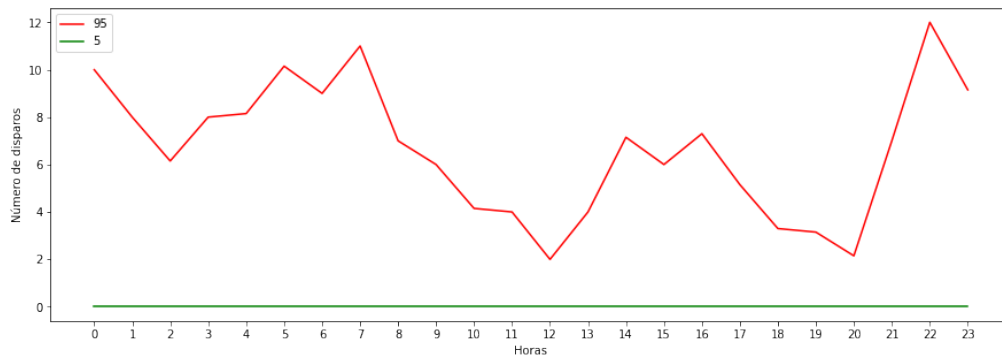


Figura 6.8: Disparos no compartimento *Quarto*. Percentil 95 e 5

### 6.1.3 Probabilidade de Ocupação

O cálculo da probabilidade de, num determinado compartimento, existir a presença de um habitante, foi outro mecanismo utilizado para estudar os padrões de movimento e de ocupação. Para aferir a probabilidade de ocupação do compartimento numa hora ( $h$ ), procedeu-se ao cálculo do número de disparos nessa hora, superior a um valor fixo, dividindo pelo número total de dias, conforme apresentado na Equação 5.2.

Após o cálculo da probabilidade, hora a hora, verificou-se novamente a existência de semelhança entre o gráfico de ocorrências (Gráfico 6.1) e o gráfico de probabilidades (Gráfico 6.9). O compartimento *cozinha*, por ser o compartimento com maior número de movimentações ao longo de todo o dia, consegue atingir valores altos

quando calculamos a probabilidade de estar ocupado, nomeadamente nos intervalos de tempo das 8 às 13 horas e das 18 às 20 horas (Figura 6.9).

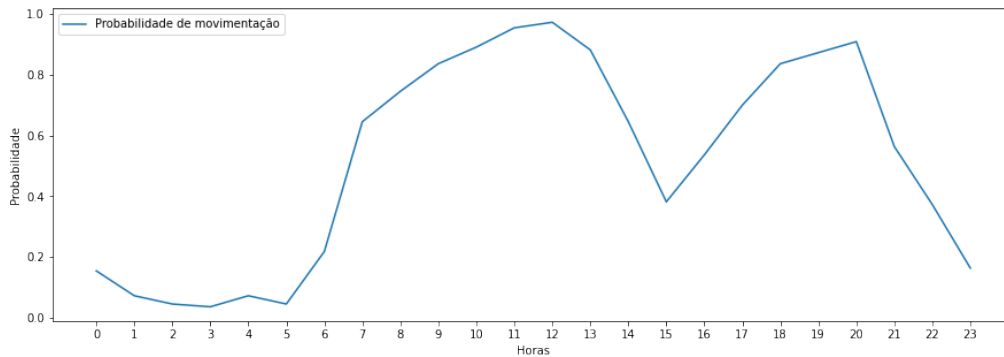


Figura 6.9: Probabilidade de ocupação da *Cozinha*

Já no compartimento *sala*, a confirmação da presença junto ao final do dia fica visível quando calculado o valor da probabilidade, atingindo valores altos no período entre as 21 e as 0 horas (Figura 6.10). Este resultado é particularmente bom quando confrontado com os valores obtidos com a mesma análise no compartimento *cozinha* (Figura 6.9).

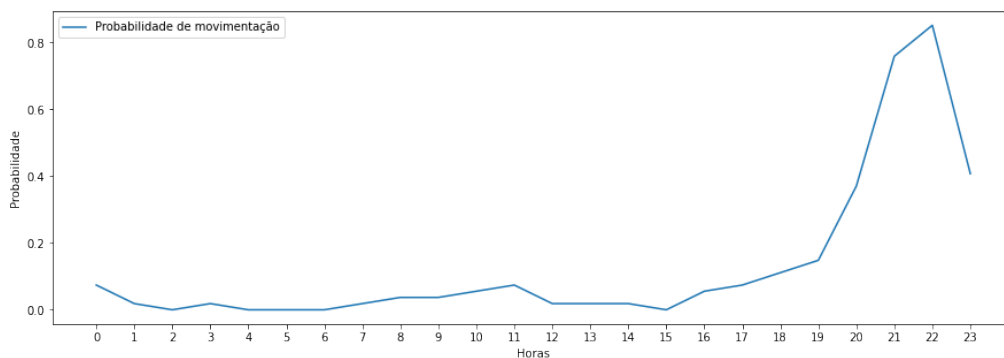


Figura 6.10: Probabilidade de ocupação da *Sala*

O compartimento *casa de banho* demonstra valores muito baixos no que diz respeito à probabilidade de existir presença, essencialmente justificado pelo propósito deste compartimento e pelo tipo de actividades de duração curta que aqui se realizam (Figura 6.11). Ainda assim, os valores mais altos observados, nos períodos entre as 8 e 9 horas e as 21 e 22 horas, estão em concordância com os padrões anteriormente identificados (Figura 6.7).

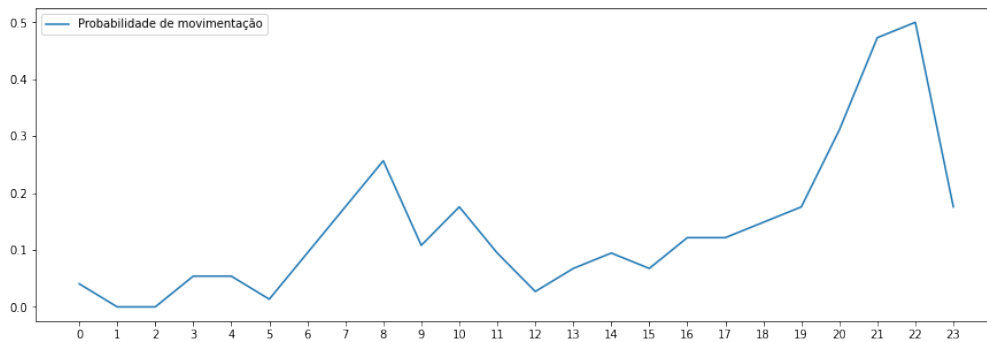


Figura 6.11: Probabilidade de ocupação da *Casa de Banho*

Os valores da probabilidade de ocupação no compartimento *quarto* (Figura 6.12) revelam mais uma vez a convergência com os padrões anteriormente identificados, nomeadamente com os valores mais altos a sobressaírem durante o período da noite entre as 0 e as 6 horas, o período da sesta entre as 14 e as 15 horas e o período final do dia por volta das 22 horas.

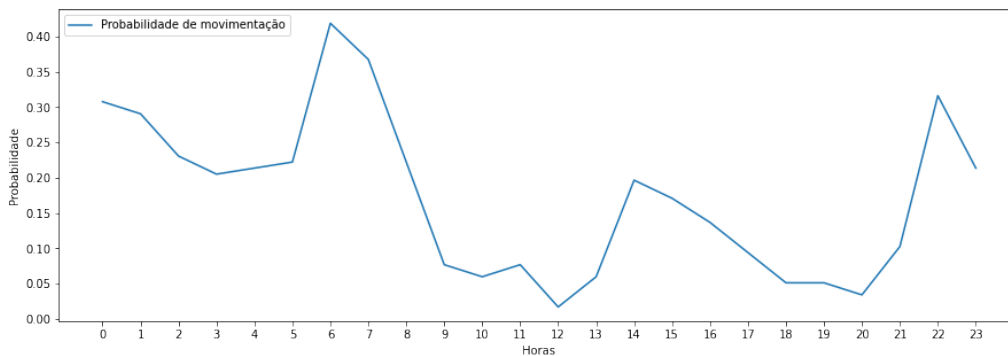


Figura 6.12: Probabilidade de ocupação do *Quarto*

Ao observarmos a probabilidade de ocupação do espaço, em todos os compartimentos, conseguimos facilmente observar também os próprios padrões de movimento, conforme apresentado na Figura 6.13.

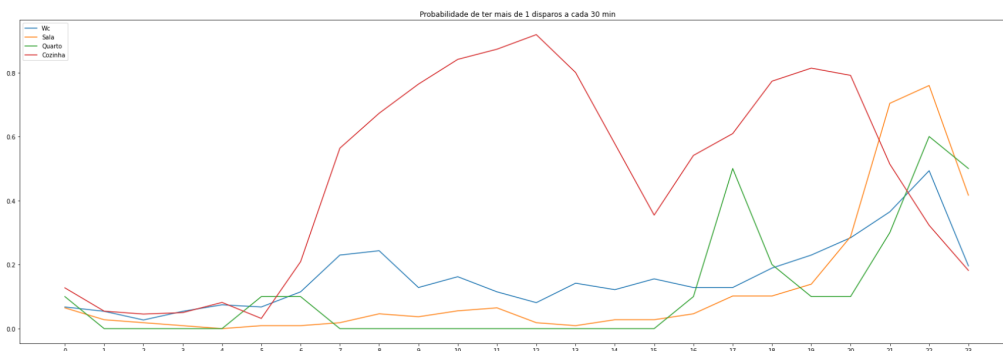


Figura 6.13: Probabilidade de ocupação em todos os compartimentos

## 6.2 Deteção de *outliers*

Foi utilizado o algoritmo de *clustering isolation forest* para calcular a proximidade dos pontos em função da combinação dos valores observados das *features*, nomeadamente a hora, o dia, o mês e o acontecimento, em cada um dos compartimentos. A cada ponto é atribuído um *score*, e podemos considerar como *outlier* todos os pontos com *score* inferior a zero.

Após o treino do modelo, com os parâmetros definidos, nomeadamente o *contamination* com o valor de 0.1 (10%) conseguimos obter resultados interessantes no que diz respeito à identificação de *outliers*. Este modelo identificou um total de 3.6% de *outliers* no conjunto de teste fornecido. É visível na Figura 6.14 a baixa quantidade de valores inferiores a zero, ou seja aqueles que são considerados *outliers*. Paralelamente à análise gráfica, podemos ainda identificar *outliers* quando o valor de *score* é inferior a zero e o valor da *prediction* é  $-1$ . A classificação de registos como *outliers* pode ainda ser condicionada pela criação de um intervalo de valores que represente os *outliers*.

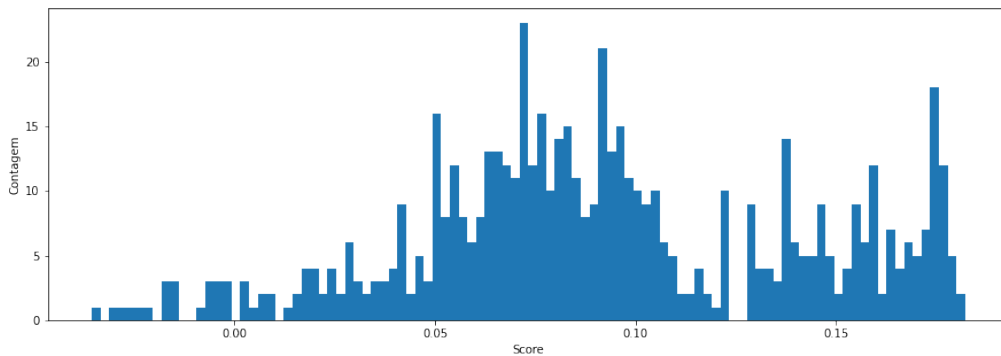


Figura 6.14: *Histograma Outliers*

Quando analisamos os *outliers* identificados, conseguimos constatar que foram identificadas situações de divergência, verificadas através da análise dos registos, a hora em que aconteceu e o local onde se detectaram movimentações (Figura 6.15).

dia_semana	hora	cozinha	sala	quarto	casa_de_banho	score	prediction
4	23.0	1	0	1	0	-0.004654	-1
6	2.0	0	1	1	1	-0.016904	-1
6	14.0	0	1	1	0	-0.015443	-1
6	23.0	0	0	1	0	-0.021297	-1
7	5.0	0	1	1	0	-0.027471	-1
7	17.0	1	0	1	1	-0.025279	-1
7	23.0	1	1	1	0	-0.014824	-1
5	5.0	0	1	0	1	-0.006145	-1
6	4.0	0	1	0	1	-0.009018	-1
7	0.0	1	0	1	0	-0.002024	-1
7	5.0	0	1	1	1	-0.035581	-1
6	23.0	1	0	1	0	-0.001213	-1
5	5.0	0	1	1	1	-0.003214	-1

Figura 6.15: Listagem dos registos *outliers*

De forma a validar a capacidade do algoritmo de deteção de *outliers*, simulámos o cálculo do *score* com dados de cenários criados para o efeito.

Se submetermos à análise do modelo, um registo com os valores: *dia da semana* 1 (segunda-feira), hora 12.0, ocupação *cozinha* 1, ocupação *sala* 0, ocupação *quarto* 0, ocupação *casa de banho* 0, obtemos um *score* positivo, portanto a identificação de um registo não *outlier* (Figura 6.16).

Dia	Hora	Cozinha	Sala	Quarto	Casa de banho	score	prediction
1.0	12.0	1.0	0.0	0.0	0.0	0.149743	1

Figura 6.16: Registo "normal"

O cenário anterior pode ser validado se o confrontarmos com a análise de padrões feita na secção 6.1.1. Já se o evento de teste representar o mesmo dia e a mesma hora mas sem a presença na *cozinha* e com ocupação da *sala* e *quarto*, o algoritmo identifica um *outlier* (Figura 6.17).

Dia	Hora	Cozinha	Sala	Quarto	Casa de banho	score	prediction
1.0	12.0	0.0	1.0	1.0	0.0	-0.041352	-1

Figura 6.17: *Outlier* detectado

Se tivermos em conta a distribuição dos valores de *score*, que foram calculados (Gráfico 6.14), percebemos que grande parte dos registos tem um *score* superior a 0, o que representa consistência nos dados. O facto de encontrarmos uma grande quantidade de valores com o mesmo *score* ou bastante próximos, percebemos que existe padronização.

### 6.3 Discussão de resultados

O reconhecimento de padrões foi alcançado em todos os compartimentos. A identificação dos padrões de movimento foi confirmada através das diferentes análises efectuadas, valores acumulados (secção 6.1.1), percentis (secção 6.1.2), probabilidade (secção 6.1.3) e ainda pelo próprio mecanismo de detecção de *outliers* (secção 6.2). Ainda que o dataset possa conter irregularidades nos registos o modelo de *machine learning* treinado, demonstrou capacidade de adaptação aos padrões identificados. A adaptação do modelo confirma-se por apenas 3.6% dos registos de teste terem sido identificados como *outliers*.

A identificação de *outliers* e o estudo para a identificação dos padrões, são dois processos distintos. Ainda assim quando analisados em conjunto, garantem a correta interpretação da informação, nomeadamente através da validação da informação dos diferentes estudos.

## Capítulo 7

# Conclusão e trabalho futuro

Este projeto permitiu identificar padrões de movimento e detectar *outliers*. O conjunto de objetivos propostos no início da realização do projeto foi atingido, desde logo pela possibilidade de obter os dados da rede de sensores, em seguida o reconhecimento dos padrões de movimento nos diferentes compartimentos e, por fim, graças ao modelo treinado, a identificação de situações anómalas.

Desta forma reconhece-se a necessidade de dar continuidade ao projeto não apenas pela sua pertinência mas também pelas diferentes oportunidades de melhoria identificadas ao longo do desenvolvimento.

O projeto contou com vários desafios, alguns deles apenas reconhecidos durante a execução. Por exemplo, a situação de, na habitação estarem dois habitantes e não apenas um, o que, por si só, causa ruído nos dados, não sendo possível determinar com exatidão os padrões de movimento. Ainda assim, os padrões foram observados, o que nos leva a concluir que o casal partilha maioritariamente as rotinas diárias. No que aos sensores diz respeito, a pequena quantidade de sensores e o tipo de sensores utilizados podem ser fatores relevantes na qualidade dos resultados alcançados. A duplicação do número de sensores será provavelmente uma opção dado que um maior número de registos será criado em cada um dos compartimentos, contribuindo para um *dataset* ainda mais rico. Ao nível da capacidade de processamento local, ainda que não seja exigida grande capacidade, verificaram-se por vezes anomalias no funcionamento da unidade, o que obrigou à criação de um mecanismo de *reboot* diário, nomeadamente durante o período da noite. Ainda que o *reboot* seja rápido pode ter-se corrido o risco de não registar algumas observações.

A detecção de *outliers* representa um processo complexo e exigente. Será relevante continuar este estudo fazendo uso de diferentes algoritmos e comparando naturalmente os resultados dos mesmos. Dada a possibilidade de existência de falsos positivos num momento inicial de treino do algoritmo não serão de excluir a utilização de outros dispositivos, como câmaras de vídeo, de forma a poder validar situações de emergência e evitar deslocções desnecessárias.

Outros sensores, não intrusivos, tais como sensores para as janelas e portas, podem ajudar a melhor entender os padrões, podendo estes fornecer até informação relevante para a interpretação dos movimentos na habitação.

Será relevante, como trabalho futuro, explorar o treino e a implementação de algoritmos de ML capazes de distinguir situações anómalas que devam ser notificadas de outras, bem como, o desenvolvimento dos respetivos mecanismos de notificação dos cuidadores.

Como trabalho futuro fica a forte vontade e o forte interesse pela continuação da implementação da solução e procura numa melhoria constante.

# Referências

- [AAL, 2021] AAL (2021). Healthy ageing research projects delivering success | AAL projects. [Citado na página 9]
- [Ageing@Coimbra, 2020] Ageing@Coimbra (2020). Monitorização remota da saúde | ageing@coimbra. <https://ageingcoimbra.pt> - Última visita: 12/3/2022. [Citado na página 2]
- [alzheimerportugal, 2020] alzheimerportugal (2020). Casos de demência em portugal vão mais do que duplicar nos próximos 30 anos. [Citado na página 3]
- [Amazon, 2022] Amazon (2022). Smart home devices and systems. <https://www.amazon.com/Smart-Home/>. [Citado na página 19]
- [Amián et al., 2021] Amián, J. G., Alarcón, D., Fernández-Portero, C., and Sánchez-Medina, J. A. (2021). Aging living at home: Residential satisfaction among active older adults based on the perceived home model. *International Journal of Environmental Research and Public Health*, 18. [Citado na página 2]
- [ANCS, 2020] ANCS (2020). Projeto 10 mil vidas - ancs. <https://www.ancs.pt/projeto-10-mil-vidas—ancs.html> - Última visita: 12/3/2022. [Citado na página 2]
- [Andò et al., 2015] Andò, B., Cristian, S. B., Lombardo, O., and Marletta, V. (2015). Smart homecare technology and telehealth dovepress smart homecare technologies for the visually impaired: recent advances. [Citado na página 9]
- [Apple, 2022] Apple (2022). ios home apple. <https://www.apple.com/ios/home/>. [Citado na página 19]
- [Aqara, 2021] Aqara (2021). Motion sensor quick start guide. [Citado nas páginas vii e 18]
- [Bhowmik et al., 2022] Bhowmik, T., Mojumder, R., Ghosh, D., and Banerjee, I. (2022). Iot based smart home-health monitoring system using dempster-shafer evidence theory for pandemic situation. In *23rd International Conference on Distributed Computing and Networking, ICDCN 2022*, page 260–265, New York, NY, USA. Association for Computing Machinery. [Citado na página 16]

- [Breiman and Cutler, 2005] Breiman, L. and Cutler, A. (2005). Random forest classification description. [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm). [Citado na página 11]
- [Casaccia et al., 2019] Casaccia, S., Bevilacqua, R., Scalise, L., Revel, G. M., Astell, A. J., Spinsante, S., and Rossi, L. (2019). Assistive sensor-based technology driven self-management for building resilience among people with early stage cognitive impairment. *2019 IEEE International Symposium on Measurements and Networking, M and N 2019 - Proceedings*. [Citado na página 9]
- [CASAS, 2013] CASAS (2013). Center for advanced studies in adaptive systems. <http://casas.wsu.edu/>. [Citado na página 12]
- [Chen et al., 2018] Chen, W. L., Chen, L. B., Chang, W. J., and Tang, J. J. (2018). An iot-based elderly behavioral difference warning system. *Proceedings of 4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018*, pages 308–309. [Citado na página 9]
- [Choi et al., 2020] Choi, Y. K., Thompson, H. J., and Demiris, G. (2020). Use of an internet-of-things smart home system for healthy aging in older adults in residential settings: Pilot feasibility study. *JMIR aging*, 3(2):e21964. [Citado na página 16]
- [Cipriani et al., 2020] Cipriani, G., Danti, S., Picchi, L., Nuti, A., and Fiorino, M. D. (2020). Daily functioning and dementia. *Dementia & Neuropsychologia*, 14. [Citado na página 3]
- [Cloud, 2022] Cloud, S. (2022). Home shelly cloud. <https://shelly.cloud/>. [Citado na página 19]
- [Dahmen et al., 2017] Dahmen, J., Cook, D. J., Wang, X., and Honglei, W. (2017). Smart secure homes: A survey of smart home technologies that sense, assess, and respond to security threats. *Journal of reliable intelligent environments*, 3:83. [Citado na página 16]
- [Debes et al., 2016] Debes, C., Merentitis, A., Sukhanov, S., Niessen, M., Frangiadakis, N., and Bauer, A. (2016). Monitoring activities of daily living in smart homes: Understanding human behavior. *IEEE Signal Processing Magazine*, 33(2):81–94. [Citado nas páginas vii e 17]
- [Denecke, 2021] Denecke, K. (2021). What characterizes safety of ambient assisted living technologies? [Citado na página 4]

- [Domashova and Kripak, 2021] Domashova, J. and Kripak, E. (2021). Identification of non-typical international transactions on bank cards of individuals using machine learning methods. *Procedia Computer Science*, 190:178–183. [Citado na página 28]
- [Domoticz, 2022] Domoticz (2022). Domoticz. [https://www.domoticz.com/wiki/Main\\_Page](https://www.domoticz.com/wiki/Main_Page). [Citado na página 20]
- [Doshi et al., 2019] Doshi, J., Patel, T., and Bharti, S. K. (2019). Smart farming using iot, a solution for optimally monitoring farming conditions. *Procedia Computer Science*, 160:746–751. [Citado na página 15]
- [Edemekong et al., 2021] Edemekong, P. F., Bomgaars, D. L., Sukumaran, S., and Levy, S. B. (2021). Activities of daily living. *Encyclopedia of the Neurological Sciences*, pages 47–48. [Citado na página 3]
- [El-Azab, 2021] El-Azab, R. (2021). Smart homes: potentials and challenges. *Clean Energy*, 5:302–315. [Citado na página 16]
- [Elhadi et al., 2021] Elhadi, S., Marzak, A., and Sael, N. (2021). Iot short-range network protocols: Analytical study and operating models. *EAI Endorsed Transactions on Internet of Things*, 7(25). [Citado na página 15]
- [Faouzi and Janati, 2020] Faouzi, J. and Janati, H. (2020). pyts: A python package for time series classification. *J. Mach. Learn. Res.*, 21:46–1. [Citado na página 28]
- [Fiorini et al., 2017] Fiorini, L., Cavallo, F., Dario, P., Eavis, A., and Caleb-Solly, P. (2017). Unsupervised machine learning for developing personalised behaviour models using activity data. *Sensors*, 17(5). [Citado na página 11]
- [Fonseca, 2020] Fonseca, A. M. (2020). Aging in place, envelhecimento em casa e na comunidade em portugal. [Citado na página 2]
- [Forbes et al., 2020] Forbes, G., Massie, S., and Craw, S. (2020). Fall prediction using behavioural modelling from sensor data in smart homes. *Artificial Intelligence Review*, 53(2):1071–1091. [Citado na página 16]
- [Furszyfer Del Rio et al., 2020] Furszyfer Del Rio, D. D., Sovacool, B. K., Bergman, N., and Makuch, K. E. (2020). Critically reviewing smart home technology applications and business models in europe. *Energy Policy*, 144:111631. [Citado na página 16]
- [Ganesan et al., 2019] Ganesan, B., Gowda, T., Al-Jumaily, A., Fong, K. N., Meena, S. K., and Tong, R. K. (2019). Ambient assisted living technologies for older adults with cognitive and physical impairments: a review. *European review for medical and pharmacological sciences*, 23:10470–10481. [Citado na página 10]

- [Gazis and Katsiri, 2021] Gazis, A. and Katsiri, E. (2021). Smart home iot sensors: Principles and applications-a review of low-cost and low-power solutions. *International Journal on Engineering Technologies and Informatics*, 2(1):19–23. [Citado na página 16]
- [Ghayvat et al., 2019] Ghayvat, H., Awais, M., Pandya, S., Ren, H., Akbarzadeh, S., Mukhopadhyay, S. C., Chen, C., Gope, P., Chouhan, A., and Chen, W. (2019). Smart aging system: Uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. *Sensors (Switzerland)*, 19. [Citado na página 4]
- [Gligoric et al., 2019] Gligoric, N., Krco, S., Hakola, L., Vehmas, K., De, S., Moessner, K., Jansson, K., Polenz, I., and Kranenburg, R. V. (2019). Smarttags: Iot product passport for circular economy based on printed sensors and unique item-level identifiers. *Sensors 2019, Vol. 19, Page 586*, 19:586. [Citado na página 15]
- [Gochoo et al., 2019] Gochoo, M., Tan, T. H., Liu, S. H., Jean, F. R., Alnajjar, F. S., and Huang, S. C. (2019). Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and dcnn. *IEEE journal of biomedical and health informatics*, 23:693–702. [Citado na página 9]
- [Google, 2022] Google (2022). Control your smart home google assistant. <https://assistant.google.com/smart-home/>. [Citado na página 19]
- [Home-Assistant, 2021] Home-Assistant (2021). Awaken your home. <https://www.home-assistant.io/>. [Citado na página 20]
- [Hunter, 2007] Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95. [Citado na página 22]
- [ioBroker, 2022] ioBroker (2022). Welcome. <https://www.iobroker.net/>. [Citado na página 20]
- [Jeedom, 2022] Jeedom (2022). Jeedom. <https://www.jeedom.com/site/en/index.html>. [Citado na página 20]
- [Jo et al., 2021] Jo, T. H., Ma, J. H., and Cha, S. H. (2021). Elderly perception on the internet of things-based integrated smart-home system. *Sensors*, 21(4). [Citado na página 18]
- [Kearney et al., 2018] Kearney, K. T., Presenza, D., Sacca, F., and Wright, P. (2018). Key challenges for developing a socially assistive robotic (sar) solution for the health sector. *IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD*, 2018-September. [Citado na página 4]

- [Kelly et al., 2019] Kelly, D., Delaney, D., and Nag, A. (2019). Fall detection with privacy as standard. *2019 International Conference on High Performance Computing and Simulation, HPCS 2019*, pages 63–69. [Citado na página 9]
- [Kurnianingsih et al., 2019] Kurnianingsih, Nugroho, L. E., Widyawan, Lazuardi, L., and Prabuwno, A. S. (2019). Detection of anomalous vital sign of elderly using hybrid k-means clustering and isolation forest. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2018-October:913–918. [Citado na página 41]
- [Kuzminykh et al., 2017] Kuzminykh, I., Sniurov, A., and Carlsson, A. (2017). Testing of communication range in zigbee technology. *2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2017 - Proceedings*, pages 133–136. [Citado na página 18]
- [Kwon et al., 2021] Kwon, L.-N., Yang, D.-H., Hwang, M.-G., Lim, S.-J., Kim, Y.-K., Kim, J.-G., Cho, K.-H., Chun, H.-W., and Park, K.-W. (2021). Automated classification of normal control and early-stage dementia based on activities of daily living (adl) data acquired from smart home environment. *International Journal of Environmental Research and Public Health*, 18(24). [Citado nas páginas 4, 9, 10 e 11]
- [Larson, 2021] Larson, E. B. (2021). Evaluation of cognitive impairment and dementia. *Obtido de www.uptodate.com*. [Citado na página 3]
- [Li et al., 2021] Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21. [Citado na página 10]
- [Liu et al., 2008] Liu, F. T., Ting, K. M., and Zhou, Z. H. (2008). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 413–422. [Citado nas páginas 29 e 41]
- [Magalhães and Pereira, 2020] Magalhães, F. M. and Pereira, M. L. (2020). *Regulamento Geral de Proteção de Dados: Manual Prático 3ª Edição Revista e Ampliada*. Vida Economica Editorial. [Citado na página 6]
- [Mahmood et al., 2020] Mahmood, Y., Kama, N., Azmi, A., and Ya’acob, S. (2020). An iot based home automation integrated approach: impact on society in sustainable development perspective. *Int. J. Adv. Comput. Sci. Appl*, 11(1):240–250. [Citado na página 17]

- [Maskeliunas et al., 2019] Maskeliunas, R., Damaševičius, R., and Segal, S. (2019). A review of internet of things technologies for ambient assisted living environments. *Future Internet 2019, Vol. 11, Page 259*, 11:259. [Citado na página 15]
- [Maskeliūnas et al., 2019] Maskeliūnas, R., Damaševičius, R., and Segal, S. (2019). A review of internet of things technologies for ambient assisted living environments. *Future Internet*, 11(12). [Citado na página 4]
- [Mertens et al., 2021] Mertens, M., Debar, G., Davis, J., Devriendt, E., Milisen, K., Tournoy, J., Croonenborghs, T., and Vanrumste, B. (2021). Motion sensor-based detection of outlier days supporting continuous health assessment for single older adults. *Sensors*, 21(18). [Citado na página 11]
- [Miljkovic, 2017] Miljkovic, D. (2017). Brief review of self-organizing maps. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017 - Proceedings*, pages 1061–1066. [Citado na página 12]
- [MongoDB, 2022] MongoDB (2022). Mongoddb atlas: Cloud document database. <https://www.mongodb.com/cloud/atlas/>. [Citado nas páginas 24 e 27]
- [Monteriù et al., 2018] Monteriù, A., Prist, M. R., Frontoni, E., Longhi, S., Pietroni, F., Casaccia, S., Scalise, L., Cenci, A., Romeo, L., Berta, R., Pescosolido, L., Orlandi, G., and Revel, G. M. (2018). A smart sensing architecture for domestic monitoring: Methodological approach and experimental validation. *Sensors 2018, Vol. 18, Page 2310*, 18:2310. [Citado na página 9]
- [Nižetić et al., 2020] Nižetić, S., Šolić, P., de-Ipiña González-de Artaza, D. L., and Patrono, L. (2020). Internet of things (iot): Opportunities, issues and challenges towards a smart and sustainable future. *Journal of Cleaner Production*, 274:122877. [Citado na página 16]
- [OECD, 2021] OECD (2021). *Health at a Glance 2021*. OCDE. [Citado na página 2]
- [Olmedo-Aguirre et al., 2022] Olmedo-Aguirre, J. O., Reyes-Campos, J., Alor-Hernández, G., Machorro-Cano, I., Rodríguez-Mazahua, L., and Sánchez-Cervantes, J. L. (2022). Remote healthcare for elderly people using wearables: A review. *Biosensors 2022, Vol. 12, Page 73*, 12:73. [Citado na página 9]
- [OpenHab, 2021] OpenHab (2021). empowering the smart home. <https://www.openhab.org/>. [Citado nas páginas 20 e 35]
- [Panico et al., 2020] Panico, F., Cordasco, G., Vogel, C., Trojano, L., and Esposito, A. (2020). Ethical issues in assistive ambient living technologies for ageing well. *Multimedia Tools and Applications*, 79(47):36077–36089. [Citado na página 24]

- [Parocha and Macabebe, 2019] Parocha, R. C. and Macabebe, E. Q. B. (2019). Implementation of home automation system using openhab framework for heterogeneous iot devices. In *2019 IEEE International Conference on Internet of Things and Intelligence System (IoT&IS)*, pages 67–73. [Citado nas páginas vii e 26]
- [Pinto et al., 2019] Pinto, M., Pereira, M., Raposo, D., Simoes, M., and Castelo-Branco, M. (2019). Gameaal-an aal solution based on gamification and machine learning techniques. *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2019*. [Citado na página 4]
- [Python, 2022] Python (2022). Welcome to python.org. <https://www.python.org/>. [Citado nas páginas 21, 32 e 36]
- [Rajeswari et al., 2019] Rajeswari, N., Nachammai, S., Jemima, P. E., and Rajeswari, A. M. (2019). Unexpected health issues prediction in medical data using apriori rare based outlier detection method. *Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019*. [Citado na página 28]
- [RaspberryPI, 2022] RaspberryPI (2022). Teach, learn, and make with raspberry pi. <https://www.raspberrypi.org/>. [Citado na página 25]
- [Rawassizadeh et al., 2019] Rawassizadeh, R., Dobbins, C., Akbari, M., and Pazvani, M. (2019). Indexing multivariate mobile data through spatio-temporal event detection and clustering. *Sensors*, 19(3). [Citado na página 38]
- [Rezvani et al., 2021] Rezvani, R., Kouchaki, S., Nilforooshan, R., Sharp, D. J., and Barnaghi, P. (2021). Semi-supervised learning for identifying the likelihood of agitation in people with dementia. [Citado na página 10]
- [Sahu et al., 2021] Sahu, D., Pradhan, B., Khasnobish, A., Verma, S., Kim, D., and Pal, K. (2021). The internet of things in geriatric healthcare. *Journal of healthcare engineering*, 2021:6611366. [Citado na página 9]
- [Sandeepa et al., 2020] Sandeepa, C., Moremada, C., Dissanayaka, N., Gamage, T., and Liyanage, M. (2020). An emergency situation detection system for ambient assisted living. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. [Citado na página 4]
- [Selicato et al., 2021] Selicato, L., Esposito, F., Gargano, G., Vegliante, M. C., Opinto, G., Zaccaria, G. M., Ciavarella, S., Guarini, A., and Buono, N. D. (2021). A new ensemble method for detecting anomalies in gene expression matrices. *Mathematics 2021, Vol. 9, Page 882*, 9:882. [Citado na página 28]

- [Shah et al., 2019] Shah, A. S., Nasir, H., Fayaz, M., Lajis, A., and Shah, A. (2019). A review on energy consumption optimization techniques in iot based smart building environments. *Information 2019, Vol. 10, Page 108*, 10:108. [Citado na página 15]
- [Sharma et al., 2021] Sharma, A., Sharma, S., and Gupta, D. (2021). A review of sensors and their application in internet of things (iot). *International Journal of Computer Applications*, 975:8887. [Citado na página 16]
- [sklearn, 2022] sklearn (2022). sklearn.cluster.kmeans — scikit-learn 1.0.2 documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>. [Citado na página 12]
- [Sovacool and Furszyfer Del Rio, 2020] Sovacool, B. K. and Furszyfer Del Rio, D. D. (2020). Smart home technologies in europe: A critical review of concepts, benefits, risks and policies. *Renewable and Sustainable Energy Reviews*, 120:109663. [Citado nas páginas vii, 20 e 21]
- [Tornero-Quinones et al., 2020] Tornero-Quinones, I., Saez-Padilla, J., Espina Diaz, A., Abad Robles, M. T., and Sierra Robles, A. (2020). Functional ability, frailty and risk of falls in the elderly: Relations with autonomy in daily living. *International Journal of Environmental Research and Public Health*, 17(3). [Citado na página 2]
- [Wang et al., 2019] Wang, Z.-M., Song, G.-H., and Gao, C. (2019). An isolation-based distributed outlier detection framework using nearest neighbor ensembles for wireless sensor networks. *IEEE Access*, 7:96319–96333. [Citado nas páginas 28 e 29]
- [Wink, 2022] Wink (2022). Wink a simpler smarter home. <https://www.wink.com/>. [Citado na página 20]
- [Xiaomi, 2021] Xiaomi (2021). Xiaomi life style. [Citado na página 24]
- [Xu and Pombo, 2019] Xu, L. and Pombo, N. (2019). Human behavior prediction though noninvasive and privacy-preserving internet of things (iot) assisted monitoring. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 773–777. [Citado na página 17]
- [Yahaya et al., 2021] Yahaya, S. W., Lotfi, A., and Mahmud, M. (2021). Detecting anomaly and its sources in activities of daily living. *SN Computer Science*, 2(1):1–18. [Citado na página 12]

- 
- [Yang, 2021] Yang, X.-S. (2021). Genetic algorithms. *Nature-Inspired Optimization Algorithms*, pages 91–100. [Citado na página 11]
- [Zambrano-Montenegro et al., 2019] Zambrano-Montenegro, D., Bellido-Outeiriño, F. J., García-Bermúdez, R., Flores-Arias, J. M., and Huhn, A. (2019). Advanced monitoring system for daily activity in elderly people. *2019 IEEE International Conference on Consumer Electronics, ICCE 2019*. [Citado na página 4]
- [Zantalis et al., 2019] Zantalis, F., Koulouras, G., Karabetsos, S., and Kandris, D. (2019). A review of machine learning and iot in smart transportation. *Future Internet 2019, Vol. 11, Page 94*, 11:94. [Citado na página 15]
- [Zikria et al., 2021] Zikria, Y. B., Ali, R., Afzal, M. K., and Kim, S. W. (2021). Next-generation internet of things (iot): Opportunities, challenges, and solutions. *Sensors*, 21(4). [Citado na página 17]