



Universidade do Minho

Mestrado Engenharia Informática

Sensorização e Ambiente

MONITOR DE OCUPAÇÃO E COMPORTAMENTOS EM LOCAIS PÚBLICOS

Grupo 7

PG50384 Francisco Reis Izquierdo,

PG50528 José Pedro Martins Magalhães



29 de abril de 2023

Conteúdo

1	Introdução	4
1.1	Definição do Domínio	4
1.2	Objetivos	5
1.3	Proposta de Solução	5
1.4	Estrutura do Relatório	5
2	Arquitetura	6
2.1	Ferramentas Utilizadas	6
2.1.1	Placa ESP8266	6
2.1.2	Servidor DeepNote	6
2.1.3	Locker Studio	7
2.2	Funcionamento do Sistema	7
2.2.1	Back-end	9
2.2.2	Aplicação Web	9
3	Tratamento e Exploração dos Dados	10
3.1	Tipos de Dados	10
3.2	Tratamento dos Dados	10
4	Sistema Desenvolvido	12
5	Modelos de Machine Learning	16
6	Resultados	17
7	Trabalho Futuro e Conclusão	19
7.1	Trabalho Futuro	19
7.2	Conclusão	19

Lista de Figuras

2.1	Arquitetura do Sistema	8
4.1	Interface do Sistema	13
4.2	<i>HeatMap</i>	13
4.3	<i>HeatMap</i> com intervalo de tempo = 0	14
4.4	<i>HeatMap</i> com $\sigma = 10$	14
4.5	Visualização dos Dados - Looker Studio (página 1)	15
4.6	Visualização dos Dados - Looker Studio (página 2)	15
5.1	Previsão do modelo com base na hora do dia	16
6.1	Dados recolhidos por dia da semana	17
6.2	Percentagem de utilização das zonas da biblioteca	17
6.3	Quantidade de dispositivos por hora	18
6.4	Distancias capturadas por sensor	18

Lista de Tabelas

3.1	Tabela de atributos gerais	10
3.2	Tabela de atributos gerados	11

1. Introdução

O presente relatório surge no âmbito da disciplina de Sensorização e Ambiente, pertencente ao perfil de Sistemas Inteligentes, no qual foi proposto ao grupo de trabalho a elaboração de um sistema de sensorização tirando partido da integração de sensores físicos e/ou virtuais, focando domínios emergentes como a Sensorização Móvel, a Computação Humanizada e as Cidades Inteligentes. Com isto, o âmbito deste projeto deve organizar-se em etapas distintas e progressivas. Desta forma, o intuito é desenvolver um sistema que se enquadre no contexto da recolha de dados através de sensores. Posteriormente, este sistema deve ser capaz de processar e tratar estes dados, preparando estes para uma das vertentes mencionadas e previamente selecionada. Seguidamente, tem-se o cerne do sistema, no qual este é elaborado e desenvolvido sob a forma de aplicação, por forma a ir de encontro ao previamente definido pelo grupo de trabalho. Por fim, também são explorados aspetos secundários, mas relevantes no âmbito dos domínios supracitados, no qual são explorados tanto a visualização dos dados recolhidos, bem como o desenvolvimento de modelos de *Machine Learning* que inferem relativamente aos mesmos dados.

1.1 Definição do Domínio

Nos dias de hoje é notório a presença de massas populacionais em diversos espaços públicos, sendo muitas vezes difícil de gerir estes espaços em diversos aspetos, tais como segurança, higiene e disponibilidade. Assim, a gestão destes espaços públicos traz grandes vantagens tanto para as entidades responsáveis por estes locais, na medida que permite avaliar e melhorar a qualidade do local, como também para a comunidade, na medida que permite ter um local mais seguro, higiénico e acima de tudo, sustentável do ponto de vista de taxa de ocupação. Desta forma, este âmbito insere-se no contexto das **Cidades Inteligentes**, no qual se tem como objetivo a administração, planeamento e gestão de espaços que são de domínio público. No entanto, associada a este domínio, existe a difícil tarefa de gerir e compreender a taxa ocupacional e comportamentos dos cidadãos que frequentam estes locais.

Dado o supramencionado e motivados por solucionar esta tarefa, decidiu-se desenvolver um sistema capaz de, dado um espaço público, exibir a taxa ocupacional em diferentes dias e horas do dia, bem como alguns comportamentos demonstrados pelas pessoas que frequentam esse espaço. Consequentemente, o sistema disponibilizará dados relativos à taxa de ocupação do espaço, tendo em conta o dia e hora do dia. Desta forma, qualquer pessoa que pretenda frequentar o local numa dada circunstância, poderá visualizar o estado e ocupação do mesmo, facilitando o seu processo de decisão relativamente se deve ou não frequentar o local dado o seu nível de ocupação.

Com isto, o desenvolvimento do sistema teve como base de referência de espaço público a biblioteca da Universidade do Minho, no qual esta é diariamente muito frequentada por diversos estudantes. Assim, este espaço é muitas vezes procurado por estes estudantes, com diversos propósitos, tais como sessões de estudo, sessões de leitura, pesquisa, etc. Com isto, surgem algumas inconveniências, sendo principalmente apontada a sua elevada taxa de ocupação, informação esta não disponibilizada para um estudante que se dirija para o local. Desta forma,

tenciona-se perceber e disponibilizar informação relativa à taxa de ocupação da biblioteca, por forma a resolver a inconveniência de se chegar ao local e deparar com a impossibilidade de frequentar o mesmo, por exemplo para uma sessão de estudo. Além disto, tenciona-se saber quais as atividades e áreas mais frequentadas da biblioteca, dado o dia e hora do dia por parte dos estudantes.

1.2 Objetivos

Dado a definição do domínio, pretende-se recolher dados provenientes das pessoas que frequentam a biblioteca da universidade do Minho, que de alguma forma elucidem para a ocupação da mesma, bem como quais as áreas/atividades mais procuradas por estes. Com isto e posteriormente ao tratamento e processamento dos dados, tem-se como objetivo apresentar através de um mapa de calor da planta da biblioteca, tendo como parâmetros o dia e a hora do dia, os níveis de ocupação desta, sendo esta informação disponibilizada através de uma aplicação web acessível ao público. Além disto, com o suporte dos modelos de *Machine Learning*, pretende-se prever quais os dias e horas do dia, bem como quais as áreas mais frequentadas pelos estudantes.

1.3 Proposta de Solução

Uma vez estabelecida a definição do domínio e os objetivos propostos, dado o âmbito da unidade curricular, decidiu-se apostar na recolha de dados provenientes dos dispositivos móveis dos estudantes através do uso de sensores físicos. Esta decisão deveu-se ao de que cerca de 64,4% da população mundial têm acesso à internet, sendo este acesso predominante nos jovens [5]. Com isto, estes dados serão consequentemente tratados e processados, por forma a transparecerem no mapa de calor da planta da biblioteca, consoante a hora e dia a que foram capturados, a taxa de ocupação e área da biblioteca a que dizem respeito. Além disto, tal como referido, será desenvolvido uma aplicação web, no qual qualquer estudante poderá visualizar o estado do local, bem como outros dados relativos ao mesmo. Serão também desenvolvidos mecanismo de visualização de dados, bem como previsão da taxa de ocupação através dos modelos de *Machine Learning* desenvolvidos.

1.4 Estrutura do Relatório

No que concerne à estrutura do presente relatório este encontra-se dividido em 6 capítulos, onde são abordados os temas relativos à **Introdução** onde são apresentados a definição, bem como os objetivos e como o grupo se propôs a atingi-los, seguido da **Arquitetura** onde são detalhadas as ferramentas utilizadas e como a arquitetura se organiza e o sistema funciona. Posteriormente a isto, são elucidados no capítulo **Tratamento e Exploração dos Dados** todo o tratamento e criação de novos dados a partir dos que foram previamente recolhidos, seguindo-se a apresentação da interface gráfica no capítulo **Sistema Desenvolvido**. São ainda abordados os temas relativos aos modelos de *Machine Learning* e dos resultados obtidos, nos capítulos **Modelos de Machine Learning** e **Resultados**. Por fim, são enaltecidas questões sobre trabalho futuro e respetiva conclusão, no capítulo **Trabalho Futuro e Conclusão**.

2. Arquitetura

De forma a corresponder aos objetivos e consequentes propostas de solução mencionadas, o sistema desenvolvido teve primeiramente de ter implementada toda a arquitetura que o sustenta. No âmbito geral e seguindo o fluxo de funcionamento do sistema, teremos as seguintes etapas: recolha dos dados, tratamento/processamento dos dados, análise dos dados, visualização dos dados, inferência dos dados e previsão dos dados. Com isto, para cada etapa foi necessária a utilização e/ou implementação de componentes adequados às mesmas.

2.1 Ferramentas Utilizadas

2.1.1 Placa ESP8266

Dado os objetivos propostos e o facto de serem facilmente configuráveis e programáveis, foram utilizadas duas placas **ESP8266**, no qual estas foram responsáveis pela recolha periódica dos dados, tendo sido colocadas na biblioteca, uma vez que este era o espaço de estudo. Cada placa é do tipo Arduino, sendo de baixa potência, adequada para IoT (*Internet of Things*), que facilita a ponte para Cidades Inteligentes, eliminando a necessidade de comunicação e processamento com fio. Estas foram ambas programadas e configuradas através do **IDE Arduino**, no qual os dados recolhidos por estas placas são reencaminhados pelas mesmas para um servidor. Assim, convém mencionar para cada placa algumas das especificações que foram atrativas dado o domínio do projeto:

- Capacidade Wifi (2.4 GHz band);
- 4 MB de memória flash;
- Interface micro-USB;
- Antena embutida;
- Open-source;
- Baixo consumo de energia;
- Conectividade Wifi e Bluetooth;
- Dual-core;

2.1.2 Servidor DeepNote

De forma a poder salvar e colecionar os dados recolhidos pelas placas previamente mencionadas, foi desenvolvido como colecionador de dados um servidor **Flask** na linguagem **Python** através da plataforma **DeepNote**. Este servidor corre permanentemente, no qual agrega os

dados que recebe periodicamente das duas placas em ficheiros apropriados e respetivos a cada uma, disponibilizando-os posteriormente para o seu tratamento e processamento. A escolha da utilização deste colecionador deveu-se maioritariamente à elevada quantidade de dados recolhidos pelas placas que teriam de compactuar com elevada capacidade de armazenamento dos mesmos, característica evidente por este colecionador.

2.1.3 Locker Studio

Por forma a se obter uma análise mais profunda e apelativa aos utilizadores, foi utilizado a ferramenta **Locker Studio**, no qual através dos dados previamente disponibilizados pela *back-end*, consegue obter *insights* mais precisos e relevantes. Assim, este componente é utilizado para a análise dos dados, no qual disponibiliza informações posteriormente à aplicação web, em que estas são utilizadas de forma embutida, disponibilizando a análise estatística dos dados sobre a forma de gráficos/tabelas gerados (por exemplo através de gráficos de rosca, colunas, barras e tabelas de dados).

2.2 Funcionamento do Sistema

No que concerne ao funcionamento do sistema, podemos então explicar este tendo em conta as diversas etapas do mesmo e as ferramentas utilizadas e que foram previamente descritas, além de outros componentes desenvolvidos como suporte ao sistema.

Assim, os dados são primeiramente recolhidos pelas duas placas **ESP8266** que se encontram na biblioteca, estando estas posicionadas nas extremidades opostas de uma parede em comum. Esta decisão deve-se ao facto de para a realização da implementação do sistema apenas ter sido possível obter duas placas, complicando a tarefa relativa à triangulação. Ora, o método de triangulação, diz respeito à medição e posicionamento que utiliza a análise de ângulos e distâncias entre três pontos de referência para determinar a localização precisa de um objeto ou ponto de interesse. Apesar de ser mais complexo de implementar traria mais vantagens na medida que permite obter uma maior precisão da localização das pessoas que frequentam o espaço, sendo necessário três placas para a implementação deste. Assim, dada a limitação de recursos, prosseguiu-se ao descrito acima, havendo uma abstração no que concerne à localização das pessoas, uma vez que qualquer localização calculada diz respeito a uma aproximação da mesma e não à localização exata, devido ao posicionamento e número das placas. Ainda sobre as placas, estas executam periodicamente a atividade conhecida como "sniffing", que consiste em capturar os pacotes de sondagem dos dispositivos móveis de vários utilizadores que frequentam a biblioteca, os quais são trocados por qualquer dispositivo móvel conectado à Internet.

Posteriormente enviam os dados para o servidor **Flask** da plataforma **DeepNote** que se encontra online permanentemente, sendo que este é responsável por criar e agregar a informação proveniente de cada placa num ficheiro com os dados capturados, respetivo a cada uma. Este servidor é responsável por atender os pedidos **POST** efetuados pelas placas quando pretendem enviar dados colecionados.

Posteriormente, os dados relativos a cada ficheiro de cada placa gerado pelo servidor supra-mencionado são acedidos através da **back-end** da aplicação desenvolvida, componente este desenvolvido e responsável por tratar e processar os dados. Este componente é fulcral para o

funcionamento do sistema uma vez que é o responsável por criar a informação, nomeadamente a geração do *heatmap* da planta da biblioteca que enaltece o estado de ocupação da mesma, a ser posteriormente disposta ao utilizador através da interface visual, sendo o seu funcionamento definido e detalhado em 2.2.1.

Seguidamente após os dados terem sido devidamente tratados e processados, estes encontram-se prontos para serem visualmente apresentados a um qualquer utilizador. Assim, qualquer utilizador que pretenda obter informações sobre o estado da biblioteca no que concerne a um dado dia e hora, consegue obter estas informações através da interface visual desenvolvida que recebe os dados supramencionados. Com isto, tal como expectável, foi desenvolvida a **front-end** componente responsável pela apresentação visual dos dados nesta fase e encarregado de apresentar estes visualmente através do *heatmap* da planta da biblioteca gerado pela **back-end**. Este componente é o foco principal no que concerne ao contacto direto para com o utilizador, na medida que é este com quem o utilizador contacta e atende os pedidos deste. Assim, relativo à **front-end** desenvolveu-se uma **aplicação web**, sendo o seu funcionamento definido e detalhado em 2.2.2.

Importa salientar ainda que, a **back-end** após realizar as tarefas competentes pelo tratamento e processamento dos dados disponibiliza estes ao **Locker Studio**, componente este responsável pela visualização e análise detalhada dos dados. Desta forma, o **Locker Studio** ao receber os dados provenientes da **back-end**, disponibiliza a análise estatística de forma embutida à **aplicação web**.

Ainda relativamente aos dados provenientes da **back-end**, estes são também acedidos, havendo algum tratamento e processamento mais detalhado, na tarefa da implementação dos modelos de *Machine Learning* desenvolvidos, por forma a serem criados novos dados relativos às previsões, no qual o aprofundamento da tarefas de tratamento de dados e implementação dos modelos é descrita e definida em 5.

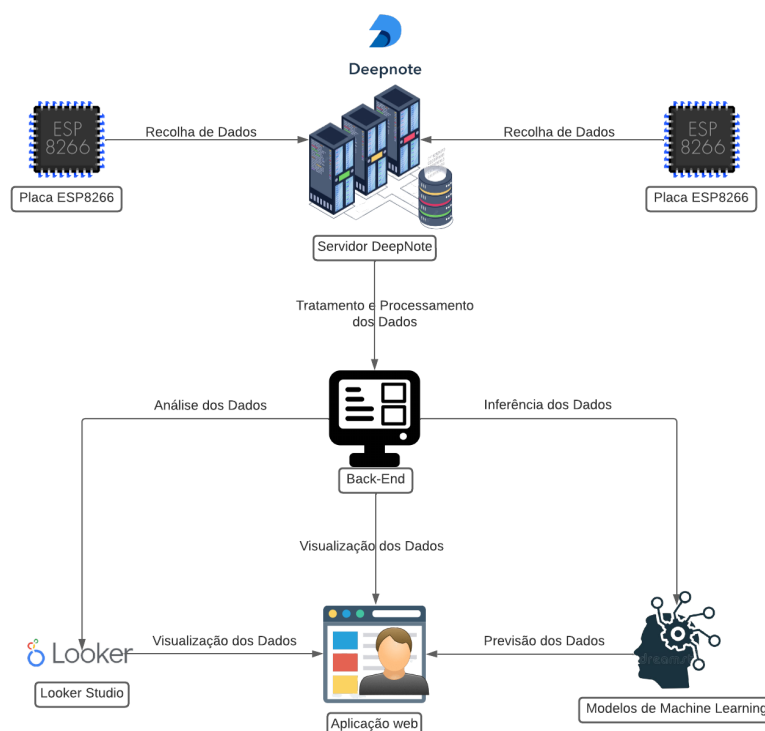


Figura 2.1: Arquitetura do Sistema

2.2.1 Back-end

Uma vez que os dados antes de poderem ser visualizados, têm de ser processados e/ou tratados, foi desenvolvida a back-end da aplicação. Esta foi desenvolvida na linguagem **Python**, cuja função é o processamento e tratamento dos dados, por forma a poder gerar o *heatmap* da planta da biblioteca associado a um dado dia e hora selecionado pelo utilizador. Ora, após ser realizado o tratamento e exploração dos dados, definidos e descritos em 3, prosegue-se à geração do *heatmap*, processo este que decorre em várias etapas distintas. Em primeiro lugar as colunas das distancias dos dispositivos a ambos os sensores são processadas gerando-se coordenadas em metros de acordo a distancia medida entre os sensores. Posteriormente, é responsável por processar os pedidos do utilizador que interage com a **aplicação web**, quando pretende visualizar o *heatmap* da biblioteca dados os parâmetros supracitados, a *back-end* transforma as coordenadas de metros para pixeis de modo a criar uma imagem de pontos que será sobreposta ao mapa da biblioteca, é também utilizado um outro parâmetro sigma como forma de suavizar o gradiente do *heatmap*. Com isto, este componente permite disponibilizar a outros componentes os dados tratados e processados de forma a estes componentes poderem também realizar as suas tarefas assim como disponibilizar à aplicação web, como resposta a um *post*, uma imagem de um *heatmap* da biblioteca numa determinada data e hora.

2.2.2 Aplicação Web

De forma ao utilizador poder visualizar e interagir com o sistema foi criada a interface gráfica, nomeadamente a **aplicação web**. Esta foi definida em **HTML** sendo responsável por disponibilizar o conteúdo gerado relativo ao *heatmap* da planta da biblioteca. Além disto, este componente é ainda responsável por mostrar toda a informação visual proveniente de outros componentes, nomeadamente da visualização dos dados relativos à análise estatística proveniente do já mencionado e detalhado **Locker Studio**. Acima de tudo, a **aplicação web** é responsável por receber os pedidos resultantes da interação por parte dos utilizadores, isto é, quando um utilizador insere informação relativa ao dia e hora do dia que deseja averiguar. Assim, após cada interação, é reencaminhado um pedido **POST** relativa à informação da interação do do utilizador (dia e hora do dia, por exemplo) para a *back-end*, no qual esta realiza as tarefas mencionadas em 2.2.1. Após obter os resultados da interação por parte da *back-end*, estes são disponibilizados posteriormente sobre a forma visual, no qual é ilucidado o novo *heatmap* gerado e que é o resultado do correspondente pedido.

3. Tratamento e Exploração dos Dados

3.1 Tipos de Dados

Dado o propósito do sistema, convém salientar e definir o tipo de dados recolhidos por este. Assim, tal como referido os dados do sistema são recolhidos periodicamente através das placas definidas em 2.1.1, no qual estes constituem informação relativa aos pacotes de *probing* dos dispositivos móveis dos diversos utilizadores que frequentam a biblioteca. Assim, periodicamente, as placas realizam o processo denominado como *sniffing*, captando estes pacotes que são trocados por quaisquer dispositivos móveis com ligação à Internet. Com isto, cada placa encaminha os dados recolhidos para o servidor, sendo posteriormente armazenados num ficheiro CSV respetivo a cada placa, no qual para cada conjunto de dados, este contém informação relativa aos pacotes capturados e cujos atributos podem ser descritos consoante a seguinte tabela.

Atributo	Descrição
MAC Address	Endereço MAC relativo ao dispositivo móvel que emitiu o pacote de probing.
Signal Strength	Força do sinal emitido pelo dispositivo móvel, aquando da emissão do pacote de probing, captado pela placa.
Distance	Distância calculada pela placa entre o dispositivo móvel emissor do sinal e a própria placa.
Timestamp	Tempo, em milissegundos, em que o pacote foi captado pela placa.

Tabela 3.1: Tabela de atributos gerais

3.2 Tratamento dos Dados

Uma vez recolhidos os dados, estes são extraídos do servidor definido em 2.1.2, sobre o formato de dois *datasets* CSV, respetivos às duas placas. Posteriormente, seguiu-se o correspondente tratamento dos dados por forma a tirar mais partido destes e constituir mais características que servissem de apoio ao desenvolvimento tanto do sistema como da visualização dos dados.

Primeiramente são renomeados alguns atributos comuns a ambos os *datasets*, nomeadamente os atributos **Distance** e **Signal Strength**. Estes são renomeados para **d1** e **d2**, (relativos à distância da placa 1 e à distância da placa 2), e **ss1** e **ss2** (relativos à força do sinal da placa 1 e à força do sinal da placa 2), respetivamente. Seguidamente, é efetuada a conversão do atributo **Timestamp** de milissegundos para minutos.

Posto isto, prossegue-se a concatenação dos dois *datasets* tendo como critério de concatenação os atributos **MAC Address** e **Timestamp**, sendo que para este último atributo é efetuada a média ponderada, por forma a evitar algum desfazamento temporal relativo aos dados proveni-

entes do mesmo dispositivo móvel. Seguidamente, são removidos dados duplicados, por forma a evitar redundância de dados que podem vir a ser prejudiciais na visualização dos resultados.

Por fim, através dos atributos definidos na tabela 3.1, os dados são explorados extraindo destes novas informações. Assim, podemos enaltecer quais os atributos e de que forma estes foram tratados por forma a conseguir obter mais informação e novos atributos.

- **MAC Addresss:** Através do endereço MAC do dispositivo móvel, foi possível definir um novo atributo associado ao tipo do dispositivo móvel, verificando o primeiro par hexadecimal associado ao endereço;
- **Timestamp:** Através do tempo em minutos é retirada informação acerca do ano, mês, dia, dia da semana, hora e minuto do dia;
- **Distance (d1 e d2):** Através das distâncias relativa às duas placas, são calculadas as coordenadas de duas dimensões, associadas ao referencial que diz respeito à planta da biblioteca. Além disso, com estas coordenadas, é possível identificar qual a área da biblioteca a que os dados dizem respeito;

Com isto, podemos acrescentar à tabela definida anteriormente, os novos atributos, resultantes do processamento e exploração dos atributos iniciais.

Atributo	Descrição
Year	Ano em que foi recolhido o sinal pela placa.
Month	Mês em que foi recolhido o sinal pela placa.
Day	Dia em que foi recolhido o sinal pela placa.
Day of the Week	Dia da semana em que foi recolhido o sinal pela placa.
Hour	Hora do dia em que foi recolhido o sinal pela placa.
Minute	Minuto da hora do dia em que foi recolhido o sinal pela placa.
X	Coordenada x relativa ao eixo de coordenadas correspondente à planta da biblioteca.
Y	Coordenada y relativo ao eixo de coordenadas correspondente à planta da biblioteca.
Device	Tipo do dispositivo móvel associado ao sinal recolhido pela placa.
Area	Área da biblioteca associado ao sinal recolhido pela placa.

Tabela 3.2: Tabela de atributos gerados

4. Sistema Desenvolvido

Uma vez detalhado toda a arquitetura do sistema desenvolvido, bem como os componentes desenvolvidos e ferramentas utilizadas na implementação do sistema, pode-se então ilucidar para os resultados obtido de todo o sistema. Assim, ao longo deste capítulo serão abordados e ilustrados diversos segmentos resultantes dos sistema, no qual serão destacadas as interfaces visuais, seja por *default* seja pelo resultado da interação de um dado utilizador.

Posto isto, iremos ilustrar o resultado produzido pelo grupo de trabalho e que é o primeiro contacto visual que o utilizador tem com o sistema. Assim, ao abrir a página web do sistema, esta contém a planta personalizada da biblioteca da universidade do Minho, no qual são enaltecidas as diversas áreas presentes nesta, sendo que podemos enaltecer as seguintes.

- **Zonas de Estudo:** Zonas identificadas pelos ícones de cadeiras e mesas, relativas a áreas de estudo;
- **Cabines de Estudo:** Zonas identificadas pelos 3 compartimentos superiores e pelos 4 compartimentos à esquerda da planta disposta. São relativos a áreas de estudo individual;
- **Zonas de Pesquisa/Leitura:** Zonas identificadas pelos ícones de prateleiras, sendo que estas estão divididas em diversas áreas;

Tal como referido anteriormente, as zonas relativas a pesquisa/leitura estão divididas em áreas, sendo estas identificadas através da legenda e respetiva cor apresentada na planta. As áreas existentes são: **Literatura, Línguas e Arte, Estatística, Sociologia, Política e Direito, Dicionários e Enciclopédias, História, Biografias e Literatura e Filosofia, Teologia, Psicologia**. Podemos também mencionar a posição das placas na biblioteca ao serem destacadas através de dois ícones, respetivos a cada placa, na parte inferior da planta da biblioteca.

Posto isto, importa ainda salientar para os mecanismos de interação com os quais o utilizador interage. Assim, existe do lado direito uma aba (para além da legenda supramencionada), que permite recolher o input do utilizador, permitindo a este selecionar o **dia, mês, ano, hora** e **minuto**. Ainda nesta aba, é permitido ao utilizador aumentar ou diminuir a intensidade dos resultados produzidos pelo mapa de calor, através do campo **sigma**. Por fim, é também permitido ao utilizador averiguar sobre o estado de ocupação da biblioteca em intervalos de tempo, através do campo **window**.

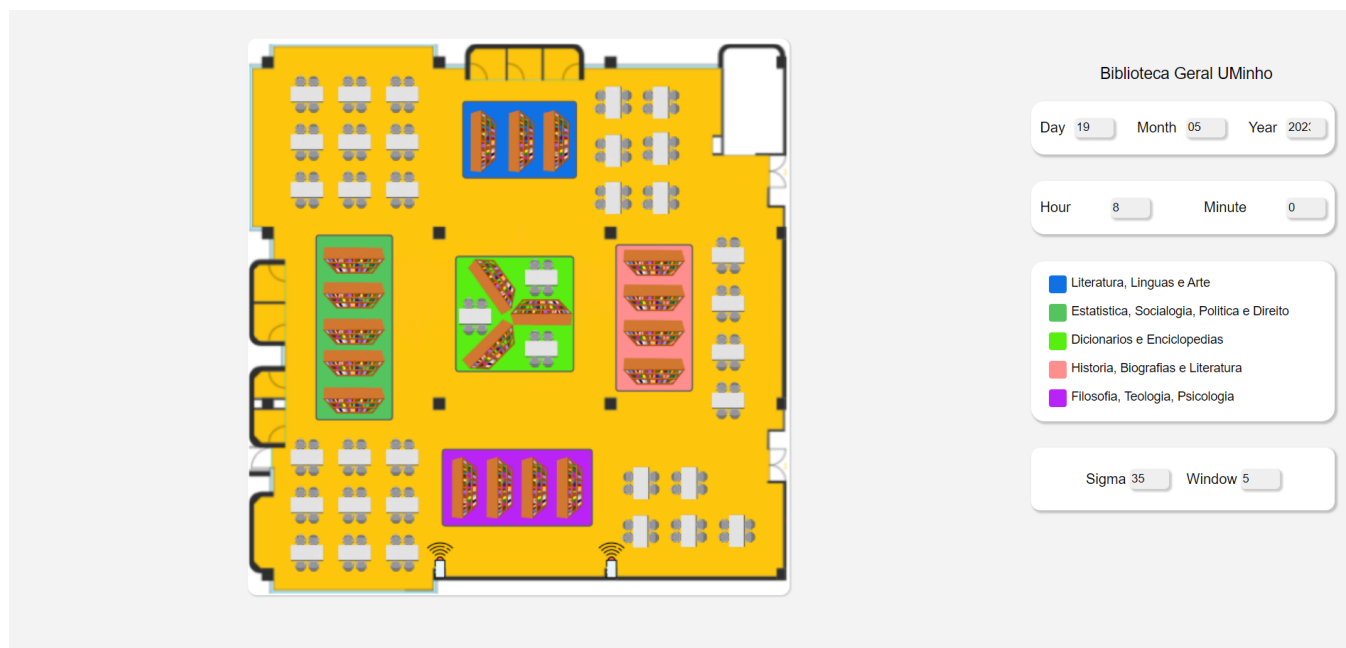


Figura 4.1: Interface do Sistema

Com isto, após o utilizador seleccionar uma data, colocando o dia, mês, ano, hora e minutos respetivos, o *heatmap* associado a essa data é produzido e ilustrado os resultados do mesmo na interface, dispondo a nova planta. Para o exemplo em questão, caso o utilizador queira perceber o estado de ocupação da biblioteca relativo a 19/05/2023 pelas 16h, ao inserir a data correspondente, irá obter o seguinte resultado.



Figura 4.2: HeatMap

Nesta nova planta é possível observar quais as zonas da biblioteca, às 16h do dia 19/05/2023 que as pessoas frequentam. De salientar para o campo **window** cujo valor está a 5, no qual está a ser tido em conta os 5 minutos antes e depois, ou seja um intervalo de 10 minutos relativos às 16h, dos resultados produzidos. Caso o utilizador pretenda observar com outro intervalo ou até sem intervalo ao fazer variar este campo obtém um resultado diferente dependendo do intervalo.

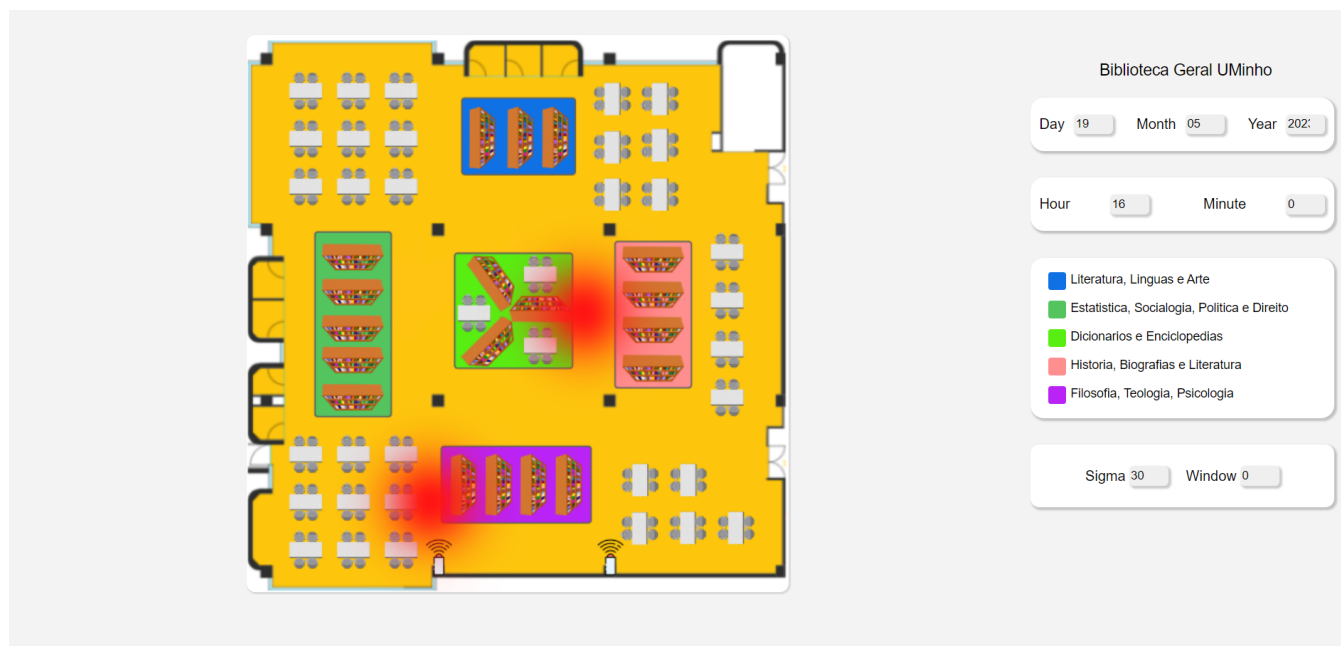


Figura 4.3: *HeatMap* com intervalo de tempo = 0

Podemos ainda salientar o campo **sigma** que também foi referido, no qual faz variar a intensidade dos resultados produzidos no *heatmap*. Assim, ao variar este o utilizador consegue ter outra perceção da disposição da ocupação da biblioteca.

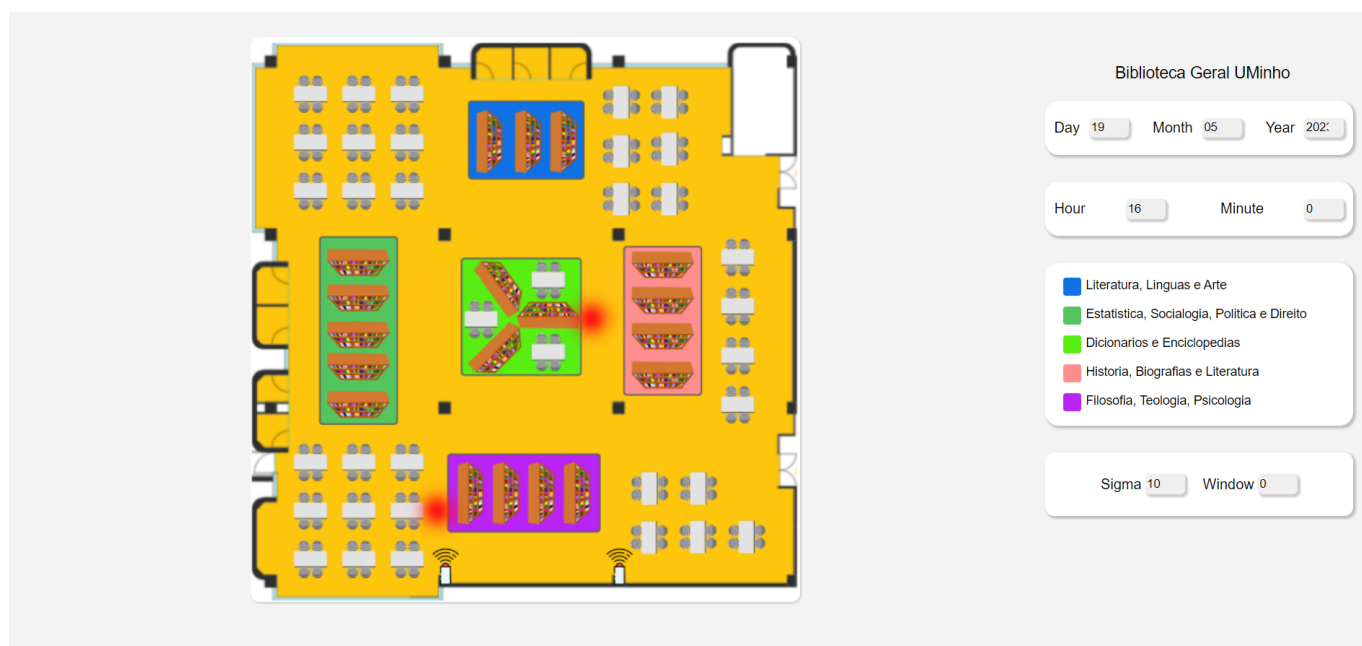


Figura 4.4: *HeatMap* com $\sigma = 10$

Outro aspeto relativo à interface do sistema que importa salientar é a visualização da análise gráfica e estatística dos dados através da utilização da ferramenta **Looker Studio** tal como referido ao longo do relatório. Esta análise é disposta ao utilizador na mesma interface web sendo esta embutida, no qual surge do lado esquerdo uma aba com duas opções relativas à disposição das diversas análises efetuadas, das quais o utilizador pode seleccionar. A correspondente análise aos resultados obtidos é detalhada e definido no tópico 6. Assim, além da visualização da planta e correspondente *heatmap* associado a um dado momento, o utilizador pode também visualizar os dados apresentados de forma gráfica.

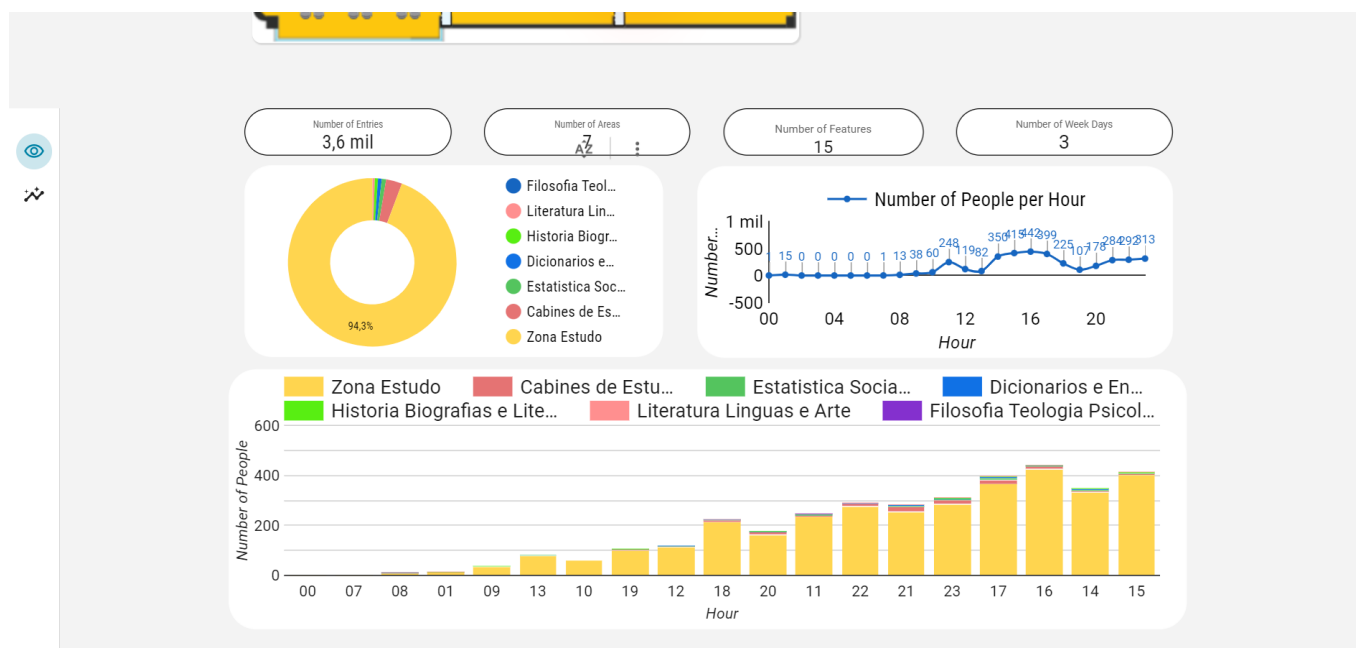


Figura 4.5: Visualização dos Dados - Looker Studio (página 1)



Figura 4.6: Visualização dos Dados - Looker Studio (página 2)

5. Modelos de Machine Learning

Dado que os utilizadores irão não só querer visualizar dados relativos ao presente e passado, uma forma de aumentar a atratividade do sistema é prever o estado do espaço, neste caso da biblioteca, tendo em conta os parâmetros previamente mencionados ao longo do relatório, isto é, a hora. Assim, com este intuito, desenvolveu-se um modelo de *Machine Learning* que têm como objetivo inferir acerca dos dados por forma a poderem realizar previsões acerca das horas do dia a biblioteca estará mais ocupada. Com isto, qualquer estudante que tenham a pretensão de utilizar a biblioteca numa determinada hora, poderá visualizar uma previsão que se aproxima do estado possível da ocupação do local.

O modelo escolhido para este problema foi uma *Random Tree Forest* com 5 *estimators* que recebe um *feature*, a hora do dia. Este modelo revelou-se bastante eficaz a prever a quantidade de dispositivos na biblioteca com base na hora do dia, obtendo-se assim os seguintes resultados. É possível observar que o nosso modelo é capaz de prever com grande precisão a quantidade de pessoas na biblioteca tendo sido capaz de modelar os vales correspondentes à hora de almoço e de jantar assim como os picos de visitantes durante a tarde.

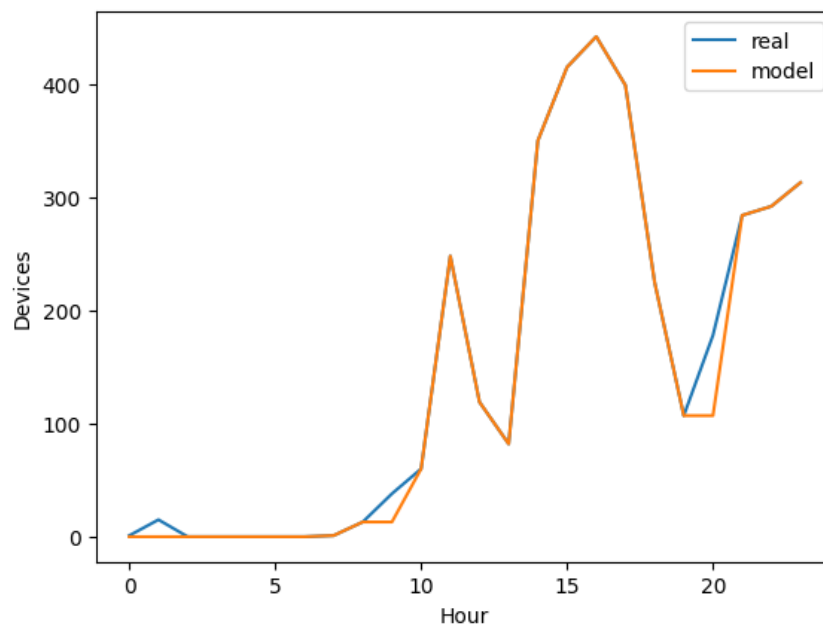


Figura 5.1: Previsão do modelo com base na hora do dia

6. Resultados

O trabalho efetuado revelou bastante informação que pode ser útil para análise e realização de melhores estratégias de planeamento na biblioteca geral da universidade. Os dados recolhidos entre o dia 18 e 20 de maio de 2023 permitem extrair conhecimento sobre as horas do dia em que a sala se encontra mais ocupada, mas também, dada a natureza deste projeto, podemos obter dados que vão muito mais além da densidade populacional neste ambiente.

Em primeiro lugar importa clarificar que os dados recolhidos dizem respeito às dadas previamente mencionadas sendo estas correspondentes a uma quinta-feira a partir das 20:00 até ao sábado seguinte às 14:30.

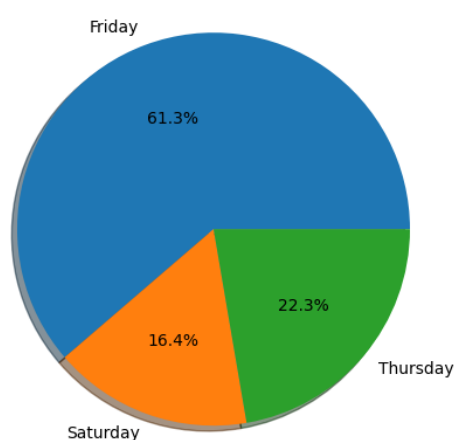


Figura 6.1: Dados recolhidos por dia da semana

Importa destacar que cerca de 94.3% das pessoas que visitam a biblioteca apenas frequentam as zonas de estudo. E apenas 3.1% das visitas se direcionam às cabines de estudo individual. Com base nestes dados e tendo em conta a margem de erro associada a este tipo de sistemas e ao desenvolvido em particular, a percentagem de pessoas que faz uso das zonas das prateleiras da biblioteca é quase nula.

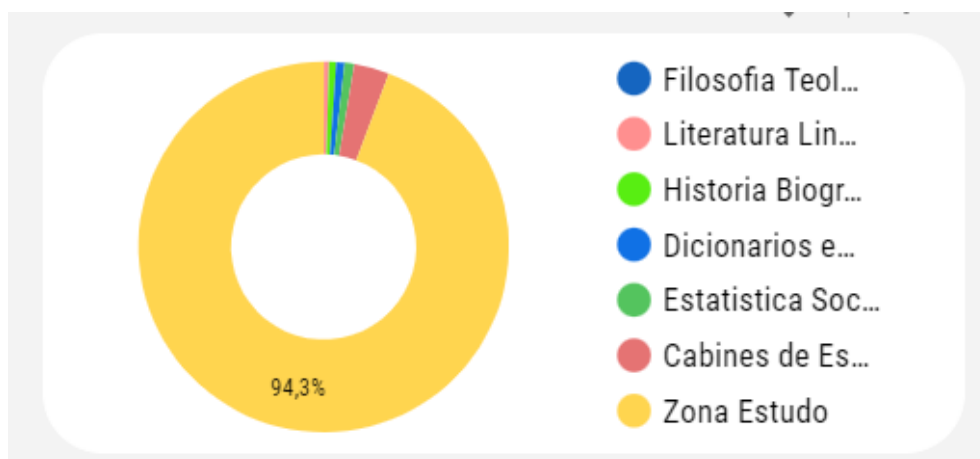


Figura 6.2: Percentagem de utilização das zonas da biblioteca

Um dos focos deste estudo baseia-se em perceber quais as horas do dia mais frequentadas pelas pessoas que visitam a biblioteca ao longo da semana. O gráfico a baixo mostra a quantidade de sinais captados durante as várias horas do dia. Tal como seria de esperar a parte da tarde e da noite são as horas mais frequentadas sendo também possível visualizar a baixa quantidade de dispositivos na hora de almoço e de jantar.

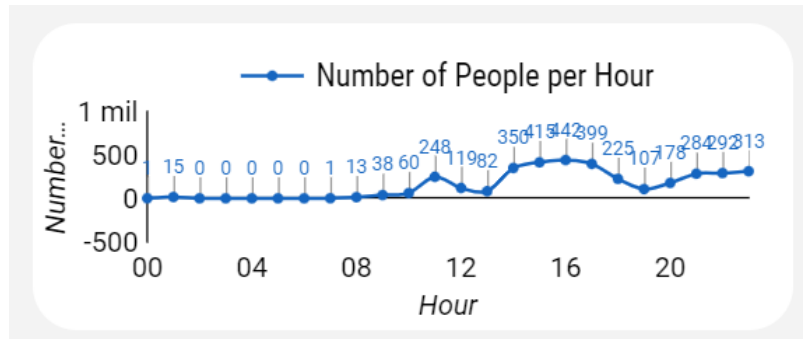


Figura 6.3: Quantidade de dispositivos por hora

No que concerne à utilização do equipamento de sensorização, podemos com algum grau de confiança e com base no gráfico a baixo perceber que o sensor 2 apresenta uma maior sensibilidade que o sensor 1 podendo isto ter afetado a posição das triangulações efetuadas assim como a quantidade de sinais captados.

Dado que os sensores captaram bastantes sinais a uma distancia superior a 24m (a largura da biblioteca) é possível que alguns dos sinais visualizados no *heatmap* assim como utilizados para a realização das estatísticas possa pertencer a dispositivos que não se encontrem na biblioteca afetando assim a sua veracidade.

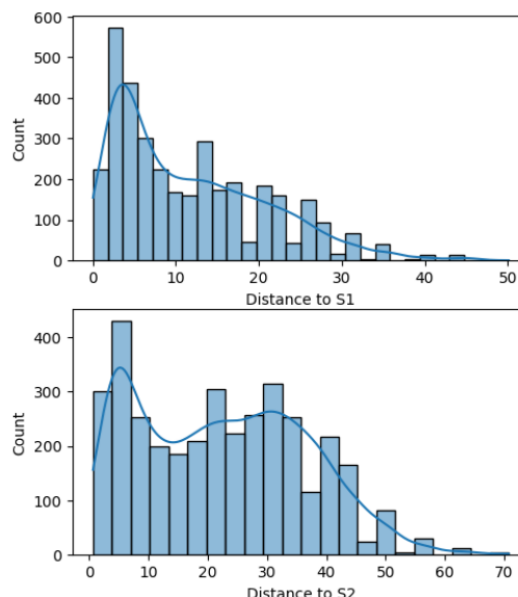


Figura 6.4: Distancias capturadas por sensor

7. Trabalho Futuro e Conclusão

7.1 Trabalho Futuro

No que concerne a trabalho futuro, existem alguns aspetos importantes a salientar. Destes destacam-se a extensão a diversos locais públicos, isto é, num trabalho futuro estender a solução desenvolvida para que esta seja versátil e adaptável a diversos espaços públicos. Desta forma, haveriam diversos locais públicos de interesse ao cidadão, no qual este poderia previamente averiguar sobre o estado de ocupação destes. Podemos também referir um dos espaços públicos que pode vir a ter mais impacto com a implementação desta solução, no qual estes são as superfícies comerciais, na medida que estas todos os dias são largamente procuradas pelo cidadão comum em "horas de ponta", em que este costuma efetuar compras. Com isto, seria possível ao cidadão previamente averiguar sobre a ocupação e quais as zonas de uma dada superfície comercial que se encontram mais preenchidas, além de facilmente poder comparar estas métricas entre diferentes superfícies comerciais e escolher aquela que tem mais disponibilidade.

Outro aspeto a ter em conta futuramente é a implementação de técnicas de recolha de dados mais precisas, nomeadamente a triangulação através de várias placas. Com isto, seria necessário um esforço monetário por forma a obter mais recursos, sendo que consequentemente e de uma forma vantajosa, a implementação da técnica supramencionada traria maior precisão relativamente às áreas frequentadas pelas pessoas. Convém ainda salientar o tópico sobre a recolha de dados, uma vez que estes, à semelhança do que aconteceu para a realização do projeto, devem ser recolhidos em conformidade com o consentimento das pessoas que frequentam o espaço público.

7.2 Conclusão

Em suma, a realização deste trabalho prático permitiu consolidar e tirar proveito de diversos conceitos abordados na unidade curricular de Sensorização e Ambiente. Destes conceitos podemos destacar o uso e implementação das placas ESP8266, por forma a recolher dados e proceder ao respetivo tratamento e processamentos dos mesmos. Outro conceito importante foi a utilização dos dados para obter insights valioso sobre o domínio em causa, isto é a taxa ocupacional e preferências dos estudantes relativamente à biblioteca da universidade do Minho. Além disto, o projeto contribui para o alargamento do conhecimento face a ferramentas de análise e visualização dos dados, que serviram de suporte para a elaboração e desenvolvimento do projeto. Por fim, importa salientar o desenvolvimento de modelos de *Machine Learning* que contribuem para o reforço e extensão da solução desenvolvida, na medida que permitem explorar e inferir acerca dos dados recolhidos.

Bibliografia

- [1] Arduino-Esp8266. Welcome to ESP8266 Arduino Core's documentation! — ESP8266 Arduino Core documentation. (n.d.).
- [2] Deepnote - Data science notebook for teams. (n.d.). Deepnote.
- [3] Arduino. (n.d.). Software.
- [4] Looker Studio Overview. (n.d.).
- [5] Kemp, S. (2023, January 26). Digital 2023: Global Overview Report. DataReportal.
- [6] UMinho Campi. (n.d.). Campi.uminho.pt
- [7] Como funcionam os receptores GPS – Trilateração vs Triangulação.(n.d.).
- [8] Scheuner, J., Mazlami, G., Schoni, D., Stephan, S., De Carli, A., Bocek, T., Stiller, B. (2016). Probr - A Generic and Passive WiFi Tracking System. 2016 IEEE 41st Conference on Local Computer Networks (LCN).
- [9] What Are Heat Maps? Guide to Heatmaps/How to Use Them. (n.d.). Hotjar.
- [10] Karanam, C. R., Belal Korany, Mostofi, Y. (2019). Tracking from one side.