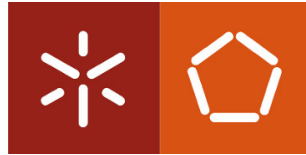


UNIVERSIDADE DO MINHO

ESCOLA DE ENGENHARIA






# Sensorização e Ambiente

Sistemas Inteligentes

Mestrado em Engenharia Informática

## Monitorização de presenças num ginásio através de sensores

Grupo 10

PG50345	PG50499	PG50779
		
Duarte Lucas	João Torres	Tiago Ribeiro

Maio, 2023

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Contextualização . . . . .	2
1.2	Principais objetivos . . . . .	2
1.3	Estrutura do relatório . . . . .	2
<b>2</b>	<b>Metodologia</b>	<b>4</b>
<b>3</b>	<b>Compreensão e estudo do problema</b>	<b>6</b>
<b>4</b>	<b>Sensorização e recolha de dados</b>	<b>7</b>
<b>5</b>	<b>Tratamento, análise e visualização de dados</b>	<b>9</b>
5.1	Tratamento e análise de dados . . . . .	9
5.2	Visualização de dados . . . . .	11
<b>6</b>	<b>Aplicação GymBox Semáforo</b>	<b>13</b>
<b>7</b>	<b>Modelos de <i>machine learning</i></b>	<b>15</b>
<b>8</b>	<b>Trabalho Futuro</b>	<b>17</b>
<b>9</b>	<b>Conclusão</b>	<b>18</b>

# 1. Introdução

Este trabalho prático surge no âmbito da unidade curricular de Sensorização e Ambiente, pertencente ao perfil de Sistemas Inteligentes do Mestrado em Engenharia Informática da Universidade do Minho. O projeto desenvolvido teve como motivação fornecer serviços personalizados e melhorar a qualidade de vida do ser humano. Recorrendo a uso de sensores, sistemas de computação em nuvem e algoritmos de *Machine Learning*, de forma a criar-se um ambiente inteligente.

Deste modo, é salientar que a gestão eficiente de filas de espera é um desafio constante em diversos setores da sociedade, desde hospitais e bancos até eventos e transporte público. A falta de um sistema adequado pode levar a longos tempos de espera, frustração dos clientes e perda de produtividade. Para ajudar a resolver esse problema, o uso de sensores tem se mostrado uma solução viável e eficaz. Diante disso, o grupo de trabalho concentrou-se nessa questão e, em colaboração com o ginásio GymBox, realizou um estudo sobre o fluxo de pessoas naquele ambiente.

## 1.1 Contextualização

A prática de atividade física é fundamental para a promoção da saúde e bem-estar, e muitas pessoas optam por frequentar ginásios como uma forma de realizar exercícios físicos. No entanto, a falta de estacionamento, a lotação das máquinas do ginásio e os horários de maior fluxo podem dificultar o acesso e a realização dos exercícios. Diante destes desafios, o grupo de trabalho pretende explorar os dados recolhidos com sensores e analisá-los de forma a gerar as melhores soluções.

## 1.2 Principais objetivos

Os principais objetivos do trabalho desenvolvido são adquirir novos conhecimentos e consolidar o conhecimento prévio sobre sensorização, abordado nas aulas teóricas e práticas. Em específico e numa versão mais prática, consolidar o conhecimento sobre o uso de *Arduino IDE*, programação de placas e utilização das mesmas para a recolha de dados. Posteriormente, estes dados serão interpretados para conclusões finais e usados para técnicas de visualização de dados em ferramentas como o *Looker Studio*.

De modo a atingir estes objetivos, podemos considerar que o trabalho divide-se em duas partes. Primeiro a programação da placa e posterior recolha de dados, e por último a fase de tratamento e visualização dos dados.

## 1.3 Estrutura do relatório

O relatório segue uma estrutura clara e organizada para apresentar os resultados do estudo de forma coerente. É dividido em nove secções principais, cada uma com um papel fundamental na análise e descrição do problema abordado.

A introdução tem como objetivo contextualizar o leitor, fornecendo informações essenciais sobre o tema e sua relevância. Dá a conhecer o que motivou este estudo e estabelece os principais objetivos do mesmo, destacando o que se pretende alcançar com a análise realizada. Esta secção é subdividida em outras três: contextualização, principais objetivos e estrutura do relatório.

A secção de metodologia descreve em detalhes a abordagem adotada no estudo. São explicadas as técnicas e métodos utilizados para realizar a análise do problema, incluindo a compreensão do problema, a sensorização e recolha dos dados e posterior tratamento dos mesmos, análise e exploração, display dos dados e conclusões. É também explicada a arquitetura utilizada.

A secção de compreensão e estudo do problema aprofunda a análise, de forma a explorar em detalhes a natureza do problema. De seguida, a sensorização e recolha de dados aborda o sensor utilizado pelo grupo, a programação efetuada no mesmo e descreve o processo de recolha de dados.

A secção de tratamento, análise e visualização de dados aborda a etapa de processamento e interpretação dos dados recolhidos, bem como a apresentação desses mesmos dados na ferramenta *Looker Studio*.

Relativamente ao capítulo da aplicação, abordamos a aplicação criada com o suporte da plataforma *Android Studio*, o semáforo do *GymBox*. De seguida, explicamos os modelos de *Machine Learning* utilizados.

Por fim, é apresentado o trabalho futuro a realizar, de forma a complementar o projeto, e a conclusão que apresenta as principais descobertas e conclusões do estudo.

## 2. Metodologia

A metodologia seguida pelo grupo de trabalho baseia-se nos seguintes passos:

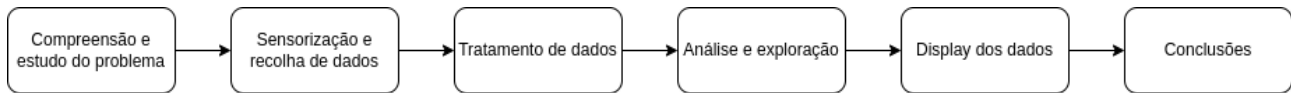


Figura 2.1: Metodologia seguida pelo grupo de trabalho

A metodologia utilizada no projeto de sensorização em um ginásio segue uma abordagem estruturada e sistemática, que pode ser dividida nas seguintes etapas:

- **Compreensão e estudo do problema:** Esta etapa envolve a identificação das necessidades e objetivos do projeto. É importante entender as questões específicas que o projeto procura resolver, como a frequência de uso do ginásio, na tentativa de melhorar a qualidade da utilização do mesmo. Também é necessário compreender o ambiente do ginásio e as variáveis que podem afetar a colheita de dados.
- **Sensorização e recolha de dados:** Nesta fase, são selecionados e instalados os sensores apropriados para recolher os dados necessários. O grupo optou por utilizar uma placa de desenvolvimento *ESP8266* para capturar informações de dispositivos dos frequentadores do ginásio.
- **Tratamento de dados:** Os dados recolhidos pelos sensores precisam de ser processados e armazenados de forma eficiente. Isto pode incluir a filtragem de dados ruidosos, a normalização de valores e a conversão de dados brutos em formatos mais úteis.
- **Análise e exploração:** Com os dados coletados e tratados, a próxima etapa é analisar e explorar os dados para extrair informações úteis e padrões. Isto pode envolver a utilização de técnicas de análise estatística, *machine learning* e visualização de dados para identificar tendências, correlações e padrões nos dados.
- **Visualização dos dados:** Após a análise e exploração, é importante apresentar os dados de forma clara e acessível. A criação de *dashboards*, gráficos e relatórios que resumam e destaquem as informações mais relevantes e interessantes obtidas a partir dos dados coletados, são formas eficientes de apresentação de dados.
- **Conclusões:** Com base nas informações obtidas nas etapas anteriores, são tiradas conclusões sobre a eficácia e os resultados do projeto de sensorização. Estas conclusões podem ser utilizadas para efetuar melhorias no ginásio, como o ajuste dos horários de funcionamento e a implementação de medidas para melhorar o conforto dos frequentadores.

Esta metodologia permite abordar de forma eficiente e eficaz os desafios e oportunidades associados à sensorização num ginásio, garantindo que as informações coletadas sejam úteis, relevantes e aplicáveis para melhorar a experiência dos frequentadores e a eficiência do espaço.

Atendendo a esta metodologia decidimos organizar uma arquitetura para o trabalho, sendo esta apresentada na figura 2.2.



Figura 2.2: Arquitetura do projeto

Inicialmente os dados são captados pelo sensor e, de seguida, enviados para o Firebase. Escolhemos utilizar esta base dados, não só porque é capaz de recebe dados em tempo real, mas também porque atende às necessidades do grupo, pois é gratuita e é capaz de armazenar o elevado volume de dados pretendido. De seguida, os dados foram enviados para as máquinas dos elementos do grupo, onde foi realizado o tratamento dos dados para que posteriormente fossem visualizados no *Looker Studio*. Esta arquitetura irá ser detalhada nas seguintes secções do relatório.

### 3. Compreensão e estudo do problema

A primeira etapa deste projeto de sensorização num ginásio é compreender e estudar o problema em questão. O objetivo desta fase é identificar as necessidades específicas do ginásio e os desafios que o projeto procura resolver. Essa compreensão é fundamental para garantir que as soluções propostas sejam relevantes, eficientes e eficazes.

Conforme já mencionado, o grupo de trabalho realizou uma reflexão sobre como as filas de espera afetam a vida quotidiana de cada cidadão, como forma de compreender e estudar o problema. Dessa forma, optou-se por analisar a afluência de um ginásio e como diversos problemas relacionados a esse tema afetam a vida das pessoas.

A análise da afluência do ginásio permitiu ao grupo de trabalho identificar os principais desafios enfrentados pelos clientes, tais como a falta de estacionamento, a lotação das máquinas e os horários de maior fluxo. Esses problemas podem levar a longos tempos de espera e afetar a experiência, causando frustração e desencorajando a prática regular de exercícios físicos. Deste modo, após uma breve conversa com o dono do ginásio *GymBox*, localizado no centro de Braga, o mesmo concordou com a exploração e compreensão deste tema no seu estabelecimento. Dando a conhecer esta proposta à administração do ginásio o grupo procura dissipar questões éticas que possam ser levantadas devido à realização deste estudo.

Ao longo desta etapa, é essencial envolver as partes interessadas relevantes, como a administração do ginásio, os funcionários e os frequentadores, para obter informações e feedback valiosos. Isto garantirá que o projeto aborda as preocupações e necessidades de todos os envolvidos e aumentará a probabilidade de sucesso e adoção das soluções propostas.

## 4. Sensorização e recolha de dados

Esta etapa corresponde à programação da placa e posterior recolha de dados. Para tal, foi adquirida a placa de desenvolvimento *WiFi ESP8266 - CP2102 NodeMCU V3 Lua*, e foi efetuada a instalação do *Arduino IDE*. Com recurso a este último, foi feito o código responsável pela captação dos dados que a placa detetou no ginásio.

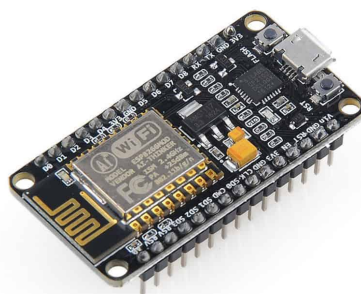


Figura 4.1: Placa de desenvolvimento WiFi ESP8266

O código desenvolvido nesta fase é um programa para o microcontrolador *ESP8266*, que tem como objetivo capturar *Probe Requests* enviados por dispositivos Wi-Fi próximos, registrar informações relevantes e enviá-las para o *Firebase*, um serviço de banco de dados em tempo real da Google.

Foram utilizadas bibliotecas que fornecem funções e recursos para a conexão a redes Wi-Fi e para interagir com o *Firebase*. São necessárias informações como o SSID e senha da rede Wi-Fi para conectar a placa à rede, bem como o endereço e a chave de autenticação da base de dados do *Firebase*.

Na função *setup*, que é executada uma vez durante a inicialização do microcontrolador, são efetuados vários procedimentos para configurar a placa. A placa é inicializada no modo combinado (AP + STA), sendo que neste modo, o *ESP8266* pode operar tanto como um AP quanto como um cliente Wi-Fi simultaneamente. De seguida, é efetuada a conexão à rede Wi-Fi, a sincronização do tempo com um servidor NTP e é inicializada a conexão com o *Firebase*.

Já a função *loop* é executada continuamente após a inicialização. Esta aguarda por 5 segundos e, em seguida, itera pelos eventos de *Probe Request* armazenados na lista de pedidos recebidos. Para cada evento, extrai o endereço MAC, o RSSI (indicador de intensidade de sinal) e a hora em que o evento ocorreu. Estas informações são enviadas para o *Firebase* em formato JSON. Após o processamento de todos os eventos na lista, ela é limpa e a função é executada novamente.

Com a programação da placa completa, foi realizada a recolha de dados. Esta fase desempenha um papel fundamental na obtenção de informações significativas e na realização de análises relevantes. Os dados foram recolhidos no contexto do ginásio *GymBox* durante uma semana, sendo o sensor colocado na entrada para o estabelecimento.

Durante o período de uma semana, desde dia 3 a 9 de maio de 2023, o ginásio operou de segunda a sexta-feira, das 8:30h às 22h, e aos sábados, das 9h às 18h. Durante essas horas de funcionamento, o microcontrolador *ESP8266* registou informações sobre a presença dos



utilizadores, como a intensidade do sinal (RSSI), o endereço MAC e a data e hora em que cada leitura foi feita, enviando estes dados para uma base de dados em tempo real fornecida pelo *Firebase*, tal como supramencionado.

Os dados foram capturados continuamente durante este período, resultando num conjunto abrangente de informações que refletem a atividade e a presença dos utilizadores no ginásio. Cada entrada no conjunto de dados corresponde a uma leitura individual realizada pelo sensor.

É importante observar que os dados coletados foram estritamente anónimos e não contêm informações pessoais identificáveis dos utilizadores. A privacidade e confidencialidade dos utilizadores foram consideradas e protegidas durante todo o processo de colheita e análise dos dados. Esta colheita foi realizada com a finalidade de realizar análises posteriormente, a fim de compreender melhor os padrões de atividade e comportamento dos utilizadores do ginásio.

## 5. Tratamento, análise e visualização de dados

### 5.1 Tratamento e análise de dados

O tratamento adequado dos dados recolhidos é essencial para garantir a qualidade e a confiabilidade das análises realizadas. Nesta secção, é discutido o processo de tratamento de dados realizado no conjunto de dados capturado no ginásio durante uma semana.

Inicialmente, os dados brutos foram obtidos a partir do sistema de monitorização e enviados para a base de dados em tempo real do *Firebase*. Esta plataforma permite descarregar os dados armazenados em formato JSON.

A raiz do objeto JSON contém uma chave denominada "dados" que representa os dados relacionados à presença dos dispositivos. Dentro da chave "dados", existem várias chaves que representam os endereços MAC dos dispositivos. Cada endereço MAC possui um objeto aninhado que contém informações sobre cada leitura específica do dispositivo. Essas informações são representadas por um identificador exclusivo e são usadas como chaves para identificar cada objeto de leitura. Dentro de cada objeto de leitura, existem as seguintes informações:

- mac: representa o endereço MAC do dispositivo.
- rssi: representa o valor do RSSI (*Received Signal Strength Indicator*) capturado pelo dispositivo.
- timestamp: representa a data e hora em que a leitura foi feita, no formato "dd-mm-yyyy HH:MM:SS".

Depois de obter o ficheiro JSON, foi iniciado o tratamento do mesmo. Primeiramente, foram retiradas todas as entradas obtidas fora do horário de funcionamento do ginásio, ou seja, apenas permaneciam no ficheiro entradas nas datas de dia 3 a 9 de maio de 2023, no período entre as 8:30h e 22h, excepto no dia 6, em que apenas são admitidas as entradas entre as 9h e 18h.

Após este tratamento inicial, relativo ao horário de funcionamento do ginásio, foi realizada a conversão deste ficheiro JSON para o formato CSV. Este novo ficheiro é constituído por quatro colunas:

- id: representa o identificador exclusivo de cada entrada de dados.
- mac: representa o endereço MAC do dispositivo.
- rssi: representa o valor do RSSI (*Received Signal Strength Indicator*) capturado pelo dispositivo.
- timestamp: representa a data e hora em que a leitura foi feita, no formato "dd-mm-yyyy HH:MM:SS".

Este CSV contém informações no mesmo formato do ficheiro JSON (mac, rssi, timestamp), porém é acrescentado um identificador para cada entrada do dataset. Após esta conversão do formato do ficheiro, foi efetuada a união dos ficheiros CSVs reunidos durante os vários dias, pois foi efetuada uma reinicialização da base de dados a cada dia.

O primeiro passo no tratamento dos dados foi a verificação e limpeza de possíveis erros ou valores ausentes. Foram realizadas análises exploratórias para identificar e tratar quaisquer inconsistências, como entradas duplicadas ou registos com campos vazios.

Ora, apenas tivemos em conta as entradas perto da porta, que era onde estava localizado o sensor. Para isso, apenas necessitamos de um *RSSI* inferior a 80. Deste modo era evitada a captação de entradas adicionais, o que iria manipular de forma negativa a nossa análise.

Para determinar este valor foram realizados testes com os telemóveis de cinco pessoas: o dos três elementos do grupo e o dos dois proprietários do ginásio. Assim, o teste dividiu-se em duas etapas. A primeira consistiu em passar pela entrada do ginásio (local onde estava localizado o sensor) e registar os *RSSI* captados em cada passagem. Verificamos que nunca ultrapassava, em média, o valor de -80.

Seguidamente, realizamos o mesmo processo, mas agora ao entrar pela porta principal do prédio, visto que o *GymBox* se localiza numa zona residencial, mais concretamente no rés do chão, onde estão constantemente a entrar e sair moradores cujos dispositivos estavam a ser captados pelo nosso sensor. Consequentemente, estavam a adulterar a veracidade dos dados recolhidos. Com isto, o grupo impôs um limite de *RSSI* nos dados recolhidos de -80, para tentar reduzir a quantidade de dados captados que não interessam a esta análise.

Durante o processo de tratamento, também foram criadas colunas adicionais para fornecer informações adicionais úteis. Por exemplo, foram criadas colunas para representar a hora do dia e o dia da semana correspondentes a cada leitura, o que permitiu uma análise mais granular e comparativa dos dados em diferentes períodos de tempo. Após o tratamento inicial dos dados, o conjunto de dados estava pronto para análises mais aprofundadas.

É importante destacar que o tratamento de dados é um processo contínuo e iterativo. À medida que novas análises são realizadas e novas descobertas são feitas, podem ser necessárias revisões e ajustes adicionais no tratamento dos dados para garantir a validade e a confiabilidade dos resultados obtidos.

## 5.2 Visualização de dados

Vamos agora abordar a visualização dos dados. Dito isto, foi utilizada a ferramenta *Looker Studio*, que permitiu a criação uma série de gráficos e métricas para avaliar de forma eficaz os dados recolhidos do ginásio.

Com o *Looker Studio*, é possível explorar visualmente as várias métricas definidas, como por exemplo, a quantidade de entradas diárias durante uma semana. Os gráficos gerados fornecem uma visão instantânea sobre o fluxo de pessoas no ginásio, permitindo identificar tendências, padrões e possíveis áreas de melhoria.

Além disso, foi implementado um servidor em *python* através da *framework Flask*, capaz de suportar as páginas de visualização e dedicado para hospedar essas *dashboards*. Posto isto, permitiu visualizar os gráficos e métricas relevantes em qualquer momento, o que facilitou a compreensão e, posteriormente, retirar conclusões com base nos dados.

De seguida, são apresentados os gráficos obtidos e a respetiva análise de resultados. De facto, estes permitem que qualquer pessoa possa entender facilmente as informações exibidas. Através destas representações visuais, é possível observar claramente o fluxo de pessoas ao longo das diferentes horas do dia e, assim, identificar períodos de maior e menor movimento.

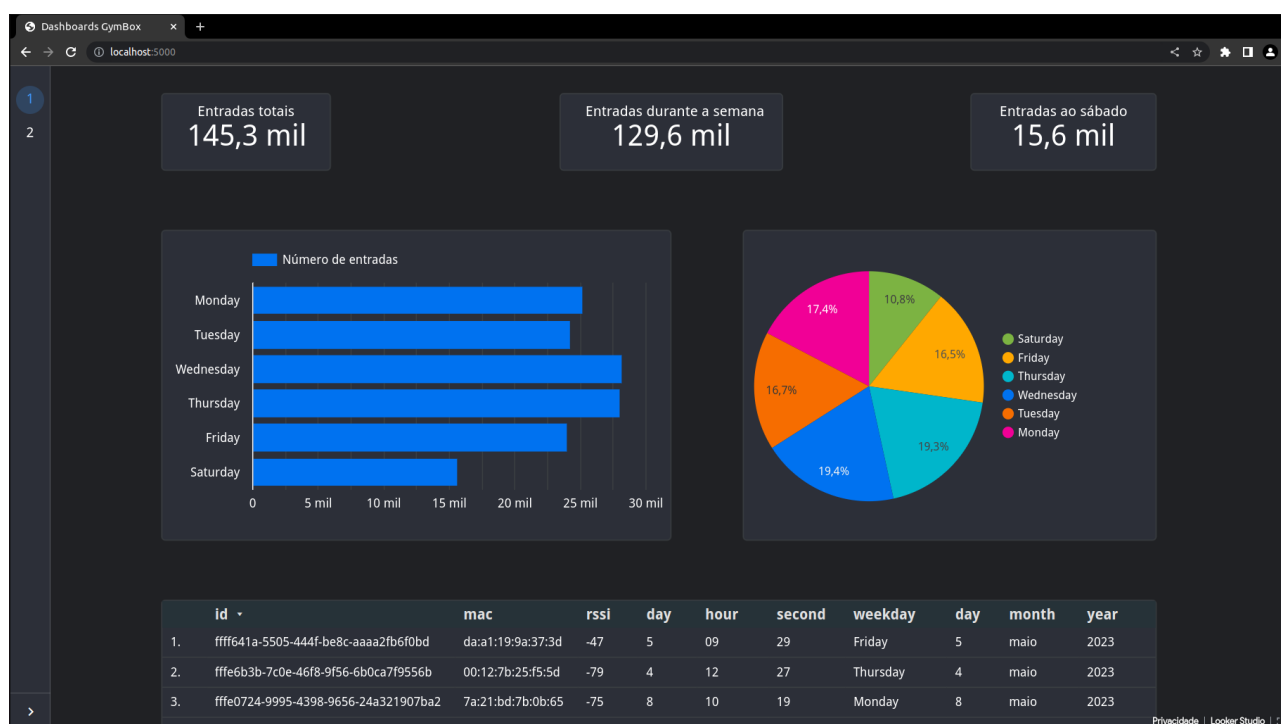


Figura 5.1: Página de visualização 1

No total, o nosso sensor captou um total de 145,3 mil entradas (pedidos Wi-Fi), sendo que 129,6 mil foram durante a semana (segunda a sexta-feira) e 15,6 mil ao sábado.

De acordo com os gráficos, podemos observar que o dia com mais afluência foi quarta-feira. A ordem decrescente de entradas foi: quarta-feira (19,4%), quinta-feira (19,3%), segunda-feira (17,4%), terça-feira (16,7%), sexta-feira (16,5%) e sábado (10,8%). De notar que o número de entradas diminui consideravelmente ao sábado, visto que o ginásio encerra às 18h.



Figura 5.2: Página de visualização 2

À segunda-feira verificou-se que os 3 momentos com maior afluência foram: primeiro às 10h, depois apresentou uma procura relativamente constante entre as 13h e as 19h (sempre à volta das 1800 entradas) e, por fim, um pico às 21h. No dia seguinte, na terça-feira, verifica-se um pico às 10h e a segunda altura do dia com mais procura é das 16h às 17h. Na quarta-feira verificou-se que os 3 momentos com mais procura foram: primeiro entre as 8:30h e as 11h, depois por volta das 15h e, por fim, perto das 19h. Durante a quinta-feira, há um elevado número de entradas entre as 9h e as 12h e na parte da tarde a procura é relativamente constante das 14 às 21 com um ligeiro aumento às 16h. À sexta-feira verifica-se um pico às 11h e na parte da tarde a procura é relativamente constante das 16h às 19h.

No sábado o ginásio abre 30 minutos mais tarde e encerra mais cedo, às 18h. A hora de maior procura do dia foi por volta das 10h. De seguida, verifica-se uma diminuição acentuada da procura até as 14h e, por fim um aumento entre as 15h e as 17h.

De forma geral, os utilizadores do *GymBox* preferem o horário da manhã para realizarem os seus treinos, mais concretamente entre as 9h e as 12h. Na parte da tarde confirma-se uma diminuição das entradas. A procura ao longo das horas deste período varia de dia para dia, logo não conseguimos retirar uma conclusão sobre a hora com maior procura. À noite, das 20h as 22h, o número de utilizadores é geralmente bastante inferior, com exceção da segunda-feira que apresentou cerca de 2100 entradas por volta das 21h. Neste último período do dia a hora com maior procura é, usualmente, as 21h.

Inicialmente, a expectativa do grupo era de que o horário de pico de utilização do GymBox estivesse próximo do final da tarde, considerando que muitas pessoas encerram suas atividades profissionais nesse período. A suposição era que os utilizadores iriam aproveitar o fim do dia para fazer exercício físico, aproveitando os períodos onde não trabalham. No entanto, após uma análise detalhada dos dados de entrada dos utilizadores, constatou-se que o horário da tarde apresenta uma diminuição de afluência, revelando que a preferência dos frequentadores está voltada principalmente para o período da manhã, sendo o pico da maior parte dos dias, durante as 10h e 11h.

## 6. Aplicação GymBox Semáforo

Após a captação dos dados e envio dos mesmos para a base de dados *Firebase* e posterior tratamento e visualização destes, o grupo optou por criar uma aplicação complementar ao trabalho realizado, sendo esta baseadaa num semáforo. Portanto, um semáforo no ginásio refere-se a um sistema de sinalização que faz uso dos dados recolhidos para indicar a disponibilidade ou ocupação das instalações. Dito isto, a aplicação foi desenvolvida através da plataforma *Android Studio*. Sendo que se trata de um semáforo, as cores escolhidas foram: verde, amarelo e vermelho.

Para definir os intervalos que iriam corresponder às cores, o grupo baseou-se nos dos dados recolhidos. Ao observar os mesmos, concluímos que as horas de maior pico tinham sempre mais de 2200 entradas e as mais calmas tinham sempre menos do que 2000. Assim definimos que a cor verde apareceria abaixo das 2000 entradas, amarelo entre 2000 e 2200 e vermelho acima das 2200.

Ora, quando um utilizador pretende ir treinar abre a aplicação e verifica a cor do sinal. Quando está verde significa que é um bom horário para ir treinar, visto que a lotação do ginásio é relativamente baixa. Quando está amarelo, a lotação está num meio termo, isto é, nem muito cheio nem muito vazio. Quando está vermelho, significa que não é muito aconselhável frequentar as instalações nesse horário, porque a lotação do ginásio está perto de estar esgotada.

De seguida, iremos apresentar algumas imagens da interface da nossa aplicação.



Figura 6.1: Aplicação fora de horas

Na figura 6.1 é possível ver a interface da aplicação quando o ginásio se encontra fechado.

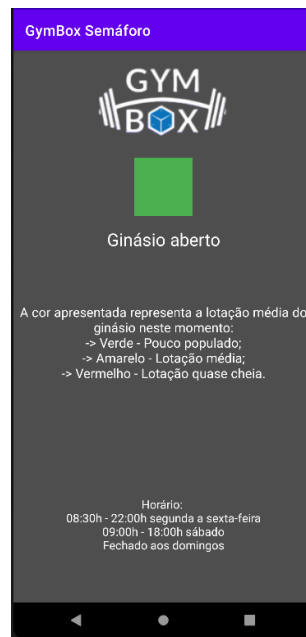


Figura 6.2: Aplicação dentro do horário com pouca lotação

A figura 6.2 mostra a interface da aplicação quando o ginásio se encontra aberto. É possível observar a cor do semáforo a verde, tal como as informações acerca do significado de cada cor. O horário continua visível, mesmo quando o ginásio se encontra aberto



Figura 6.3: Aplicação dentro do horário com muita lotação

Por fim, na figura 6.3 é possível observar a aplicação com a cor vermelha no semáforo, o que indica que o ginásio se encontra com a lotação quase esgotada.

## 7. Modelos de *machine learning*

Tal como requisitado pelo enunciado deste trabalho prático, foram concebidos e implementados modelos de *machine learning* para previsão. Tendo em conta este ponto, o grupo optou por efetuar uma previsão do número de entradas no ginásio a um certo dia. Para isso, foram implementados dois modelos de *machine learning*: *Random Forest Regressor* e *Gradient Boosting Regressor*. Foram escolhidos estes modelos de regressão, estes têm sido amplamente utilizados em tarefas de regressão e podem oferecer informações significativas para entender os padrões de tráfego e tomar decisões informadas no funcionamento e gestão de ginásios.

Assim sendo, os modelos foram treinados com os dados de segunda a quinta-feira, sendo que o objetivo é prever como se comportam as entradas durante o dia de sexta-feira. Foi excluído o dia de sábado, pois o horário de funcionamento neste dia é mais curto que nos restantes. De seguida, podemos observar os gráficos gerados, relativos a cada modelo.

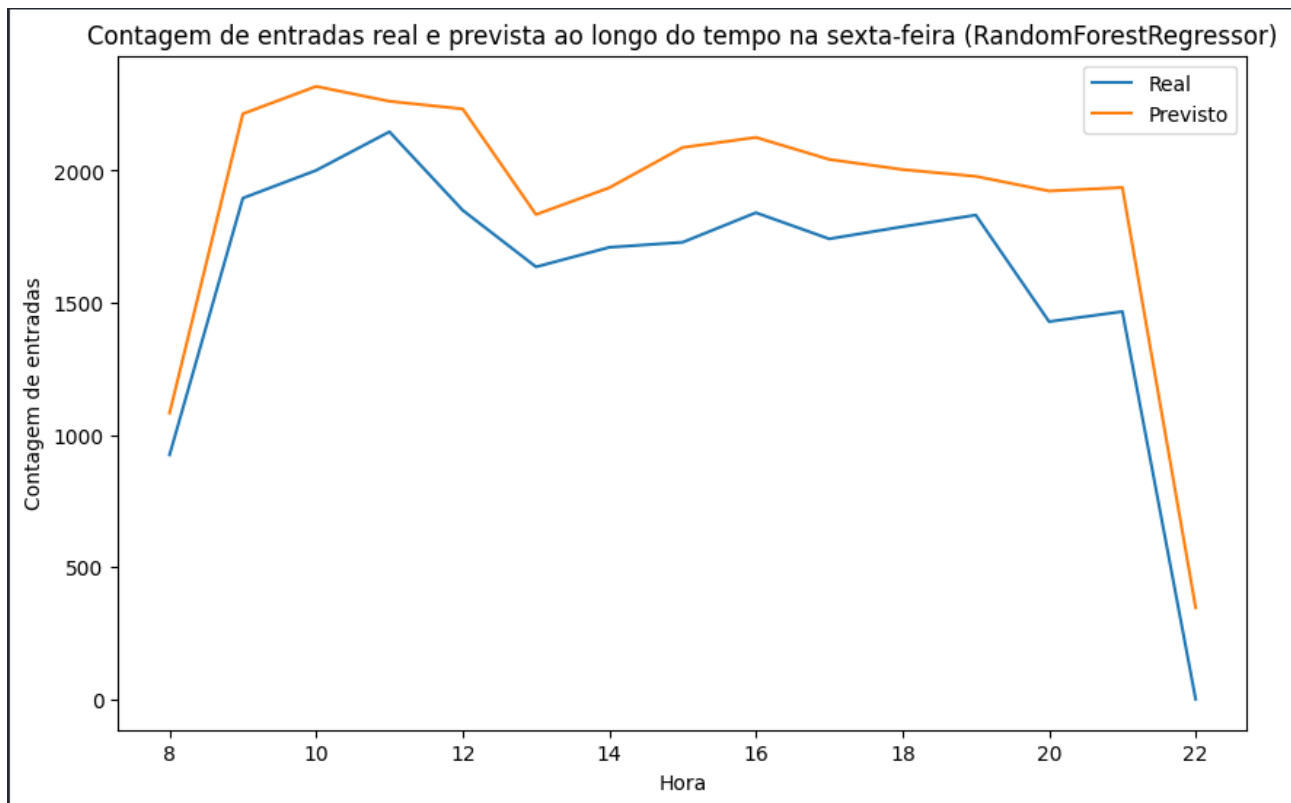


Figura 7.1: Modelo *RandomForestRegressor*



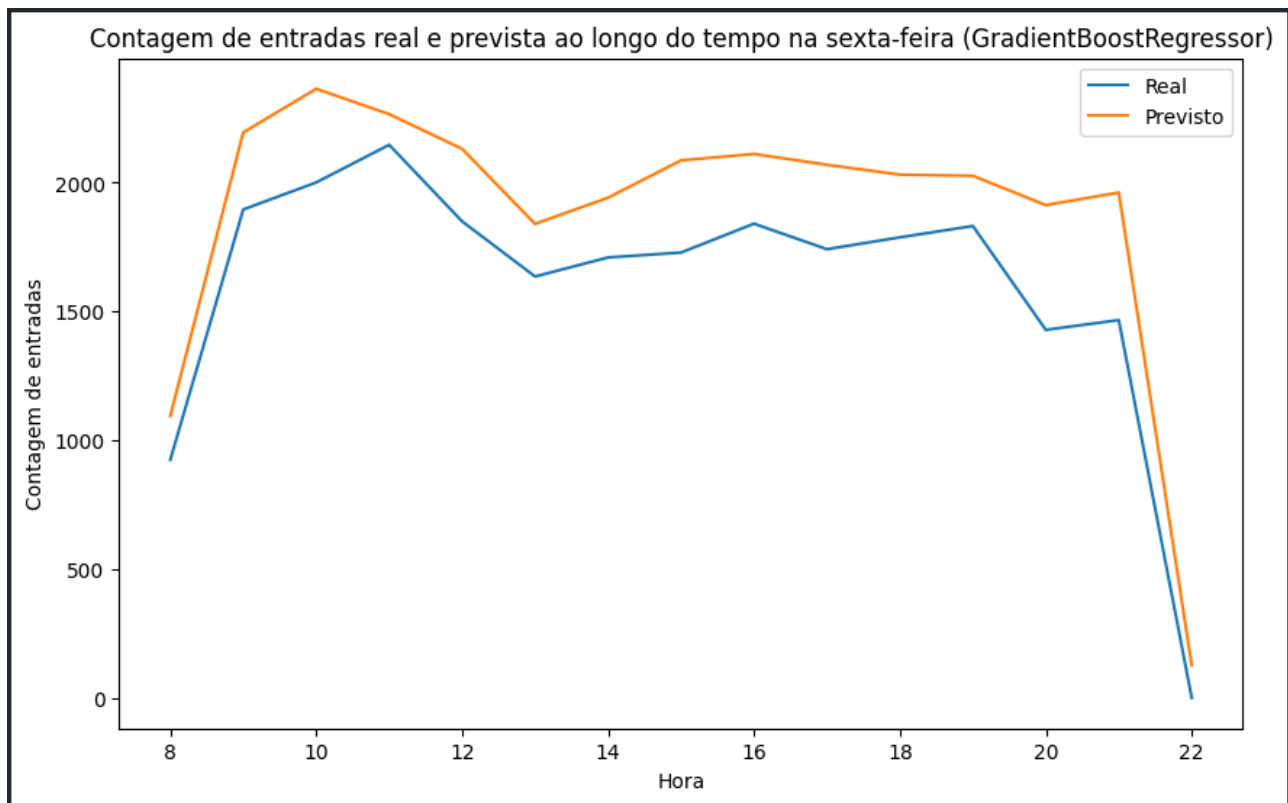


Figura 7.2: Modelo *GradientBoostRegressor*

Através da observação das figuras, concluímos que, em ambos os casos, foi capaz de detetar bem os picos de utilização do ginásio e as alturas em que o mesmo estava mais vazio.

## 8. Trabalho Futuro

No contexto do nosso projeto de monitorização das entradas num ginásio, identificamos várias oportunidades para trabalhos futuros, como objetivo de melhorar ainda mais a eficiência e a experiência dos utilizadores.

Dito isto, uma das otimizações seria a utilização de diferentes sensores, visto que os atualmente utilizados possuem limitações em termos da qualidade dos dados captados, sendo afetados pela zona residencial envolvente causando ruído nos dados.

Além disso, com o objetivo de otimizar a nossa aplicação, seria ideal que esta fosse capaz de realizar os cálculos da cor do semáforo de acordo com a ocupação do ginásio em tempo real. Isto implicaria conectar diretamente a aplicação à base de dados que recebe os dados do sensor. Com estas alterações será possível uma gestão mais dinâmica e precisa dos fluxos de utilizadores, o que irá ajudar a evitar congestionamentos.

Outra adição interessante a este projeto seria a captura de métricas adicionais. Sendo que existem diversos fatores que influenciam a frequência da ida das pessoas ao ginásio, seria uma ideia interessante, tentar incorporar a monitorização de dados meteorológicos. Ao instalar sensores de temperatura e humidade nas áreas do ginásio, tal como, monitorizar as condições meteorológicas dos dias em que foram captados os dados, poderíamos tirar conclusões acerca das condições ambientais mais adequadas para os utilizadores do ginásio.

Por fim, outra medida interessante, mas quase impossível de implementar com os meios atuais, seria a monitorização da ocupação de equipamentos. Nesse sentido, a instalação de sensores específicos em cada equipamento seria fundamental para rastrear quais estão a ser utilizados, por quanto tempo e com que frequência. Isto permitirá identificar equipamentos subutilizados ou que necessitam de manutenção, o que irá garantir a disponibilidade e a eficiência dos recursos oferecidos pelo ginásio aos utilizadores.

## 9. Conclusão

Consideramos que foi um projeto bastante interessante e desafiante, visto que nos obrigou a interligar conceitos e matéria lecionados nas aulas teóricas com a componente mais prática que nos foi instruída ao longo das aulas PLs, nomeadamente na sensorização.

O processo recolha de entradas do ginásio *GymBox* com a placa de desenvolvimento WiFi ESP8266 - CP2102 NodeMCU V3 Lua permitiu a captura eficiente das informações relevantes para análise. Os dados recolhidos durante uma semana de funcionamento do ginásio foram tratados e analisados, proporcionando *insights* importantes sobre o fluxo de utilizadores e os padrões de atividade ao longo dos diferentes dias da semana. Após o tratamento dos dados, recorremos ao *Looker Studio* para a visualização dos mesmos, o que permitiu uma compreensão clara dos padrões de entrada ao longo dos diferentes dias e horários. Os gráficos gerados facilitaram a identificação de momentos de maior e menor movimento, permitindo uma melhor gestão dos recursos do ginásio. De notar que aprendemos a trabalhar com esta nova ferramenta.

A análise crítica dos mesmos permitiu identificar tendências e padrões. Foi possível observar que o ginásio apresentou uma maior afluência durante os dias úteis, com picos de utilização às quartas-feiras. As horas de maior procura eram geralmente no período da manhã, entre as 9h e as 12h, com uma diminuição gradual durante a tarde e à noite.

A implementação de um semáforo no contexto do controlo de fluxo de entradas num ginásio é altamente benéfica e útil. Dito isto, fornece aos utilizadores informações instantâneas sobre a disponibilidade das instalações, permitindo-lhes tomar decisões informadas sobre o momento mais adequado para frequentar as mesmas. Com base nos dados recolhidos, a aplicação apresenta cores distintas para indicar a lotação atual: verde para baixa ocupação, amarelo para ocupação intermédia e vermelho para alta ocupação. Esta abordagem facilita o planeamento dos treinos dos utilizadores, pois assim conseguem evitar horários de pico e, consequentemente, têm uma experiência mais agradável no *Gymbox*.

A conceção e posterior implementação de modelos de *Machine Learning* para prever o número de entradas num dado dia trouxe uma camada adicional de análise e previsão. Utilizamos dois modelos: o *Random Forest Regressor* e o *Gradient Boosting Regressor*. Ambos foram treinados com base nos dados históricos de entrada do ginásio de segunda a quinta-feira e foram capazes de fazer previsões relativamente boas sobre o número de entradas à sexta-feira, visto que ambas as retas observadas (real e previsto) tem sempre picos semelhantes ao longo do dia.

Assim, enquanto grupo conseguimos distribuir bem o trabalho entre todos. Os membros do grupo ajudaram-se mutuamente e, de forma geral, consideramos que tivemos um aproveitamento positivo. Concluindo, este trabalho ajudou-nos a desenvolver novas aptidões e a consolidar todos os conceitos lecionados em aula, nomeadamente sobre sensorização.

# Bibliografia

- [1] Schwartz, M. (2016). Internet of Things with ESP8266. Packt Publishing Ltd.
- [2] Mesquita, J., Guimarães, D., Pereira, C., Santos, F., & Almeida, L. (2018, September). Assessing the ESP8266 WiFi module for the Internet of Things. In 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA) (Vol. 1, pp. 784-791). IEEE.
- [3] Rosli, R. S., Habaebi, M. H., & Islam, M. R. (2018, September). Characteristic analysis of received signal strength indicator from esp8266 wifi transceiver module. In 2018 7th International Conference on Computer and Communication Engineering (ICCCE) (pp. 504-507). IEEE.
- [4] Arduino, S. A. (2015). Arduino. Arduino LLC, 372.
- [5] Badamasi, Y. A. (2014, September). The working principle of an Arduino. In 2014 11th international conference on electronics, computer and computation (ICECCO) (pp. 1-4). IEEE.
- [6] Jain, A. (2015, December). A smart gym framework: Theoretical approach. In 2015 IEEE International Symposium on Nanoelectronic and Information Systems (pp. 191-196). IEEE.
- [7] Arduino IDE, <https://www.arduino.cc/en/software>.
- [8] Firebase, <https://firebase.google.com/?hl=pt>.
- [9] Looker Studio, <https://lookerstudio.google.com/>.
- [10] Android Studio, <https://developer.android.com/studio>.