

UNIVERSIDADE DO MINHO

MESTRADO EM ENGENHARIA INFORMÁTICA

Sensorização Ambiente em *Smart Cities*

Sensorização e Ambiente



Grupo 09

Gonçalo Almeida (pg47212)
Leonardo Marreiros (pg47398)
Maria Sofia Marques (pg47489)
Pedro Fernandes (pg47559)

16 de maio de 2022

Conteúdo

Lista de Figuras	i
------------------	---

Lista de Tabelas	ii
------------------	----

1 SAFit- Monitorização de saúde	1
1.1 Domínio	1
1.2 Objetivos	1
1.3 Descrição e funcionamento do sistema	1
1.4 Recolha de dados	3
1.4.1 Sensores	3
1.4.2 GoogleFit	3
1.5 Análise e tratamento dos Dados	5
1.6 Modelos de Previsão	5
1.6.1 Previsão de Atividade	5
1.6.2 Previsão de Sono	7
1.6.3 Previsão de Calorias	9
1.7 Visualização dos Dados	11
1.7.1 Visualização dos dados por ano	11
1.7.2 Visualização dos dados por mês	12
1.7.3 Visualização dos dados por dia	13
1.7.4 Visualização detalhada das atividades realizadas	14
1.7.5 Visualização detalhada do sono profundo	15
1.7.6 Visualização detalhada dos passos e calorias	16
1.7.7 Visualização das previsões	17
1.8 Análise dos Resultados	19
1.8.1 Dados organizados cronologicamente	19
1.8.2 Dados relativos às atividades	21
1.8.3 Dados relativos ao sono profundo	22
1.8.4 Dados relativos aos passos e calorias	22
1.9 Considerações Finais	22

Lista de Figuras

1 Arquitetura do sistema desenvolvido	2
---	---

2	Matriz de confusão do modelo de previsão da atividade	7
3	Diferença entre previsão e resultado verdadeiro- Percentagem de sono profundo	8
4	Diferença entre previsão e resultado verdadeiro - Minutos até primeiro sono profundo	9
5	Diferença entre previsão e resultado verdadeiro - previsão calorias	10
6	Dados por ano	11
7	Dados por mês	12
8	Dados por dia	13
9	Dados relativos às atividades	14
10	Dados relativos ao sono profundo	15
11	Dados relativos aos passos e calorias	16
12	Visualização do modelo de previsão de atividades	17
13	Visualização do modelo de previsão da percentagem de sono profundo . . .	17
14	Visualização do modelo de previsão de minutos até atingir de sono profundo	18
15	Visualização do modelo de previsão de calorias	18
16	Comparação do número de passos	20
17	Comparação da distância percorrida	20
18	Comparação de calorias queimadas	20
19	Comparação dos minutos ativos	20
20	Comparação do tempo total de sono	21
21	Comparação da frequência cardíaca	21

Lista de Tabelas

1 SAFit- Monitorização de saúde

1.1 Domínio

Atualmente, muitas pessoas usam dispositivos para acompanhar a sua saúde e condição física. Um *fitness tracker* é um dispositivo que usa sensores para rastrear a orientação, movimentação, rotação e recolhe dados como batimentos cardíacos, padrões de sono, passos e calorias gastas. O principal benefício de usar estes dispositivos é aumentar motivação a ser mais ativo. Para muitas pessoas, o exercício é mais recompensador quando têm *feedback* positivo que lhes diz que estão a fazer um bom trabalho. Outros benefícios dos *fitness trackers* incluem entender melhor a sua saúde geral, medir o progresso em relação a metas e treinar mais eficientemente.

Numa era em que estes dispositivos se tornaram extremamente populares e viáveis para uso entre a população geral, atingindo um valor de mercado de USD 36.34 bilhões americanos em 2020 ¹, esta tornou-se uma fonte de dados que vale a pena ser explorada.

1.2 Objetivos

O objetivo principal deste projeto consiste em perceber a forma como os dados recolhidos por estes dispositivos se relacionam entre si e com isto desenvolver modelos de previsão para vários fatores como: previsão do tipo de atividade que está a ser realizada baseado em batimentos cardíacos durante essa atividade, previsão da quantidade de sono profundo baseado no nível de *stress* antes de ir dormir, previsão do tempo que um sujeito demoraria a atingir o sono profundo baseado novamente no nível de *stress* antes de ir dormir, previsão das calorias gastas durante um dia baseado nos batimentos cardíacos, passos e minutos ativos, entre outros.

Para atingir estes objetivos serão necessários dados que serão recolhidos a partir de *fitness trackers* dos elementos do grupo, dados esses que são guardados na conta Google de cada um. A partir da REST API do GoogleFit esses dados serão então recolhidos.

1.3 Descrição e funcionamento do sistema

O primeiro passo é a obtenção dos dados registados pela pulseira a partir da *Google Fit* REST API. Para tal, foi necessário criar diferentes *parsers* que simplesmente transformam as chamadas de cada um dos tipos de dados de JSON para o formato posteriormente pretendido (CSV). Assim, a classe GoogleFit.java, com o auxílio destas outras classes, realiza chamadas consecutivas à API da Google Fit, e transforma a informação pretendida em diferentes ficheiros no formato CSV ao nosso gosto.

Os ficheiros CSV resultantes, são transformados num único ficheiro XLSX para serem lidos pelo PowerBI, e os datasets criados com o intuito de serem realizadas previsões são

¹<https://www.fortunebusinessinsights.com/fitness-tracker-market-103358>

passados para diversos *jupyter notebooks*.

No total foram criados 4 *notebooks*, cada um para fazer a sua própria previsão. Existem dois relacionados com o sono profundo, um que tenta prever a percentagem de sono profundo e o outro a quantidade de tempo que é necessária para entrar neste tipo de sono. Existem em seguida outros dois, agora mais ligados à atividade física, um que pretende prever a atividade que se encontra a ser realizada e um outro as calorias queimadas num dia. Numa secção seguinte irá ser realizada uma análise mais profunda a estas previsões.

Por fim, de modo a visualizar os dados, recorreu-se ao PowerBI que, lendo o ficheiro XLSX, importa toda a informação da pessoa, e permite visualizar diversas estatísticas sobre a mesma.

A arquitetura resultante encontra-se em seguida de forma ilustrativa :

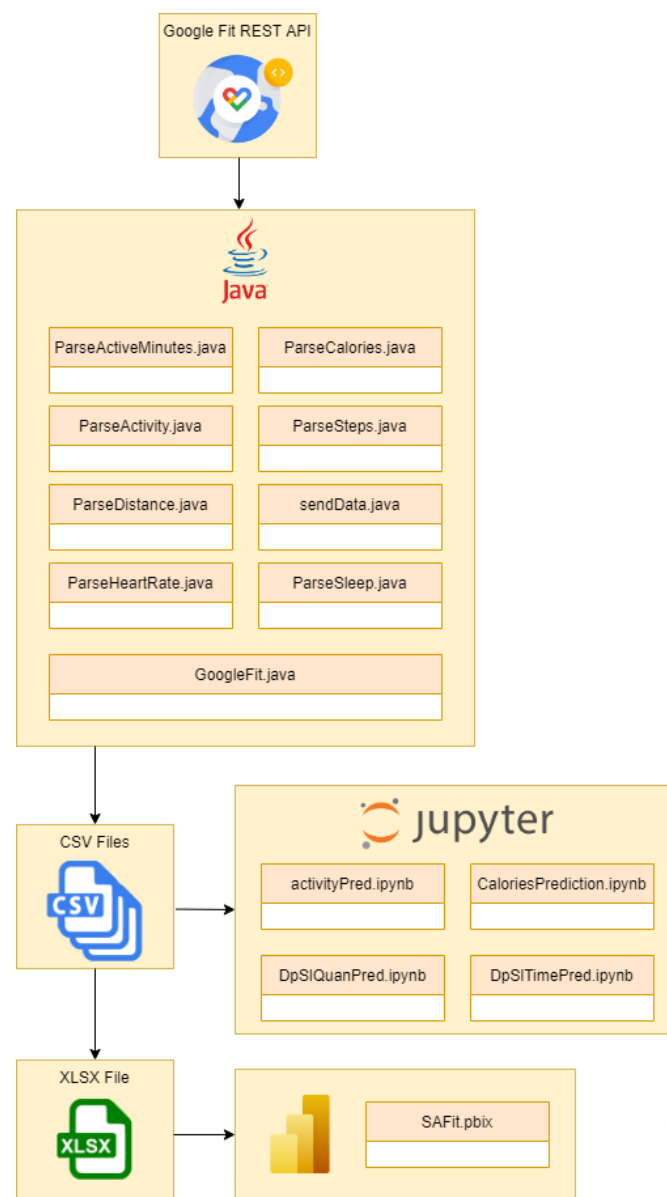


Figura 1: Arquitetura do sistema desenvolvido

1.4 Recolha de dados

Tal como referido anteriormente, cada um dos elementos do grupo utiliza um *fitness tracker*. Os dados obtidos passam pelo seguinte *pipeline*: primeiramente, os sensores presentes nos dispositivos utilizados - sensores de batimentos cardíacos, acelerómetros, giroscópios, entre outros - recolhem os dados; os dispositivos estão conetados a uma aplicação dedicada e por sua vez esta aplicação permite integração com o Google Fit; uma vez feita a integração, os dados são guardados na conta Google; finalmente, para aceder a estes dados é necessário efetuar um conjunto de passos para autorização e depois utilizar o REST API do Google Fit ².

1.4.1 Sensores

Os dados são recolhidos pelos sensores presentes nos dispositivos usados. Os sensores mais comuns e relevantes neste tipo de dispositivos são o sensor de batimentos cardíacos e o acelerómetro de três eixos. Estes dois sensores são responsáveis por recolher a maioria dos dados utilizados no projeto.

Para obter dados de sono o dispositivo monitoriza os movimentos do corpo enquanto dorme para determinar quanto tempo provavelmente passou acordado versus a dormir. Alguns dispositivos também analisam as alterações da frequência cardíaca durante o sono para estimar o tempo gasto em cada ciclo de sono.

A quantidade de passos e minutos ativos são recolhidos com auxílio do acelerómetro de três eixos que monitoriza os movimentos do corpo e permite determinar se um indivíduo está a andar, a correr, ou a fazer alguma outra atividade.

Para obter o tipo de atividade e as calorias gastas os sensores de batimentos cardíacos e acelerómetro trabalham em conjunto. Quanto mais o coração trabalha, quanto maior o esforço, mais calorias são queimadas; igualmente, uma elevada movimentação associado a um ritmo cardíaco elevado provavelmente significa algum tipo de atividade que exige maior esforço como corrida ou aeróbica.

1.4.2 GoogleFit

Uma vez integrado o GoogleFit na aplicação dedicada resta obter os dados da conta Google. Para isso utilizamos a API dedicada, *Fitness REST API*. Primeiro foi necessário solicitar um ID de cliente OAuth 2.0 que exige um conjunto de passos ³: entrar na *google cloud console* e ativar a API; configurar a tela de permissão OAuth tendo o cuidado de colocar o estado da aplicação em produção; criar uma nova credencial do tipo *OAuth Client ID*; entrar no OAuth Playground e autorizar a API de *fitness*; finalmente, obter o *access token*.

Com o *access token* é possível submeter pedidos HTTP. De seguida apresenta-se o exemplo

²<https://developers.google.com/fit/rest>

³<https://developers.google.com/fit/rest/v1/get-started>

do corpo de um pedido de tipo POST para obter dados de sono (os outros pedidos efetuados são similares):

```
1 {
2   "aggregateBy": [{
3     "dataTypeName": "com.google.sleep.segment"
4   }],
5   "bucketByTime": { "durationMillis": 86400000 },
6   "startTimeMillis": 1649256060000,
7   "endTimeMillis": 1649342420253
8 }
```

O *dataTypeName* corresponde ao tipo de dados a obter, neste caso de sono, *bucketByTime* corresponde ao intervalo em milissegundos pelo qual os dados irão ser agregados, neste caso apresenta o valor 86400000 que corresponde a um dia. Os campos *startTimeMillis* e *endTimeMillis* correspondem à data de início e de fim em milissegundos em que os dados irão ser obtidos. A seguir apresenta-se a correspondente resposta:

```
1 {
2   "bucket": [
3     {
4       "startTimeMillis": "1649256060000",
5       "endTimeMillis": "1649342420253",
6       "dataset": [
7         {
8           "dataSourceId": "derived:com.google.sleep.segment:com.google.
9 android.gms:merged",
10          "point": [
11            {
12              "startTimeNanos": "1649282040000000000",
13              "originDataSourceId": "raw:com.google.sleep.segment:com.
14 xiaomi.hm.health:GoogleFitSyncHelper- sleep segments",
15              "endTimeNanos": "1649292420000000000",
16              "value": [
17                {
18                  "mapVal": [],
19                  "intVal": 4
20                }
21              ],
22              "dataTypeName": "com.google.sleep.segment"
23            },
24            {
25              "startTimeNanos": "1649292420000000000",
26              "originDataSourceId": "raw:com.google.sleep.segment:com.
27 xiaomi.hm.health:GoogleFitSyncHelper- sleep segments",
28              "endTimeNanos": "1649294400000000000",
29              "value": [
30                {
31                  "mapVal": [],
32                  "intVal": 5
33                }
34              ],
35              "dataTypeName": "com.google.sleep.segment"
36            },
37            (...)
38          ]
39        }
40      ]
41    }
42  ]
43 }
```

Cada resposta tem uma estrutura similar. Para cada resposta é feito o parse do JSON para

obter a informação relevante. Neste caso interessaria-nos o campo *intVal* que corresponde ao tipo de sono (sono leve, sono profundo, entre outros) e os campos *startTimeNanos* e *endTimeNanos* que corresponde ao intervalo de tempo no qual se verificou esse tipo de sono, em nanosegundos.

Numa primeira iteração os dados estavam a ser obtidos até à hora em que o programa era corrido desde o dia anterior a essa mesma hora e a ser enviados para *feeds* do Adafruit. No entanto, o Adafruit apresenta um *throttle* que limitava a quantidade de dados recebidos, tornando esta solução inviável. Assim sendo, tomamos a decisão de enviar os dados para ficheiros de texto. Além disso, uma vez que um dos elementos do grupo já utilizava um *fitness tracker* há vários anos, passamos também a obter os dados históricos desde o início de 2020 até ao dia e hora em que o programa é executado.

1.5 Análise e tratamento dos Dados

À medida que fomos preparando os *datasets* reparamos que certos pedidos à API não retornavam valores e, noutros casos, retornavam valores anómalos. Uma possível causa é o facto de talvez os valores apenas serem registados pelo *fitness tracker* em certos intervalos, intervalos esses diferentes daqueles que foram pedidos. Outra razão possível é a falta de precisão dos sensores como resultado de fraco contacto com o corpo do utilizador, ou até o facto do dispositivo não estar a ser utilizado nesse momento.

Para lidar com estes problemas foi preciso tratar os dados. Este tratamento foi feito tanto nos programas Java, onde são construídos os *datasets* como nos *notebooks* onde são feitas as previsões.

Em certos casos, nos programas Java, apenas eram escritas linhas cujos valores fossem todos diferentes de nulo, noutros casos as linhas eram escritas independentemente da existência de valores nulos, passando esses campos como *null* no *dataset* e posteriormente removendo linhas com valores nulos no *notebook* correspondente.

O tratamento de valores anómalos foi feito maioritariamente nos *notebooks* e será explicado nas secções seguintes.

1.6 Modelos de Previsão

1.6.1 Previsão de Atividade

O primeiro modelo de previsão que decidimos realizar foi a previsão de uma atividade, estando incluídas nestas, não apenas as atividades físicas como também as atividades de sono (sono leve, sono profundo).

Para tal previsão, foi necessária a criação de um *dataset* adequado com os dados possíveis de ser recolhidos pela pulseira. O resultado foi um *dataset* composto pelas seguintes colunas :

- **Begin** : Data e hora do início da atividade;
- **End** : Data e hora do fim da atividade;
- **Activity** : Nome da atividade realizada;
- **HRmean** : Média da frequência cardíaca média ao longo da atividade;
- **HRmax** : Média da frequência cardíaca máxima ao longo da atividade;
- **HRmin** : Média da frequência cardíaca mínima ao longo da atividade;

A primeira coisa a ser feita em seguida foi criar uma nova coluna, **Duration**, a partir da subtração da coluna de fim e início da atividade, que representa a duração da atividade. Uma vez que eram colunas irrelevantes à previsão, em seguida foram removidas as colunas **Begin** e **End**.

Seguidamente foram retirados as linhas com valores nulos.

Dada a fraca variedade e qualidade dos dados relativos às atividades recolhidas, foi então decidido apenas prever as atividades cuja quantidade dos dados fosse relevante, ficando no final apenas com dados referentes a **Walking**, **Lightsleep**, **DeepSleep** e **Other**.

Sendo que o volume destas atividades era muito grande, e de modo a deixar o modelo balanceado recorreu-se a um *undersampling* de todos os tipos de atividades para 500, obtendo no final um modelo com cerca de 2000 linhas, uma vez que destas atividades selecionadas ainda se procedeu à remoção de *outliers*.

De modo a realizar a previsão foram então criados vários modelos com diferentes técnicas diferentes. Foi utilizada regressão linear, árvores de decisão, SVM, *Random Forests*, *XG-Boost* e Redes Neurais Artificiais.

O modelo que obteve uma melhor *accuracy* foi aquele onde foi utilizada a técnica das *Random Forests*, com um valor de 83%. Embora este valor não seja muito alto, podemos observar pela matriz de confusão da figura 2 que onde este modelo se mais engana é naquelas atividades mais semelhantes, ou seja, quando está a dormir ou acordado. Este resultado pode ter sido influenciado pela falta de precisão do sensor de frequência cardíaca da pulseira, pela pessoa proprietária dos dados ou até pela falta de outras *features* importantes para determinar as atividades que não sejam fornecidas pelo equipamento utilizado, por exemplo as variações do acelerómetro.

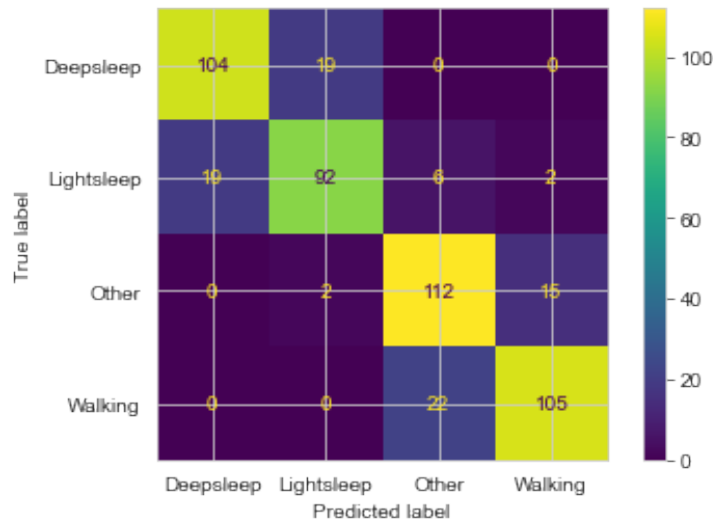


Figura 2: Matriz de confusão do modelo de previsão da atividade

1.6.2 Previsão de Sono

Um fator muito importante para a sociedade é a qualidade do sono, uma vez que este é um fator fundamental para a saúde humana. Assim sendo decidiu-se realizar duas previsões relativas a esta área, uma tentativa de prever a percentagem de sono profundo, e uma outra de prever, em minutos, o tempo até atingir o primeiro sono profundo.

Desta forma foi necessária a criação de um *dataset* cujo objetivo seria tentar prever estas métricas. Assim, embora limitados pelos sensores presentes na *mi band 4*, foram utilizadas as seguintes colunas:

- **HRmean** : Média da frequência cardíaca média, meia hora antes início do sono;
- **HRmax** : Média da frequência cardíaca máxima, meia hora antes início do sono;
- **HRmin** : Média da frequência cardíaca mínima, meia hora antes início do sono;
- **HalfHourBeforeSleep** : Hora de meia hora antes do sono;
- **DeepSleepBegin** : Hora do início do sono profundo;
- **LightSleep** : Quantidade em minutos de sono leve;
- **DeepSleep** : Quantidade em minutos de sono pesado.

Percentagem de sono profundo - Previsão 1

Para esta primeira previsão começou-se por determinar uma coluna com a duração até atingir o sono profundo (**DurationUntilDeepSleep**), desde meia hora antes de dormir(**HalfHourBeforeSleep**), até à hora do primeiro sono profundo(**DeepSleepBegin**), através da subtração destas duas colunas.

Em seguida procedeu-se à adição de uma outra coluna (**TotalSleep**) que nada mais nada menos é, do que a soma total do sono, obtido com a soma do **LightSleep** com o **DeepSleep**.

Em seguida procedeu-se à remoção de dados nulos assim como à remoção das colunas que deixaram de ter interesse, as colunas **HalfHourBeforeSleep**, **DeepSleepBegin** e **LightSleep**. Por sua vez foi criada a coluna **DeepSleepPercentage** através da divisão do **DeepSleep** pelo **TotalSleep**, multiplicado em seguida por 100.

Para finalizar a estrutura do dataset removeram-se as colunas **DeepSleep** e **TotalSleep**, obtendo-se no final um *dataset* constituído por 5 colunas :

- **HRmean** : Média da frequência cardíaca média, meia hora antes início do sono;
- **HRmax** : Média da frequência cardíaca máxima, meia hora antes início do sono;
- **HRmin** : Média da frequência cardíaca mínima, meia hora antes início do sono;
- **DurationUntilDeepSleep** : Tempo em minutos até atingir o sono profundo;
- **DeepSleepPercentage** : Percentagem de sono profundo;

Antes dos dados serem usados nos modelos, estes ainda sofreram uma remoção dos *outliers* relativos às frequências cardíacas obtidas.

Para realizar a previsão criaram-se diversos modelos, utilizando técnicas diferentes. Foi utilizada regressão linear, árvores de decisão, SVM, *Random Forests*, Redes Neurais simples e XGBoost.

O modelo que obteve melhor resultado foi aquele onde foi utilizada o SVM, com um MAE de 5.78. Visto o modelo representar uma percentagem, este valor não é mau. Na figura abaixo podemos ver as diferenças entre os valores reais e os valores previstos do modelo escolhido.

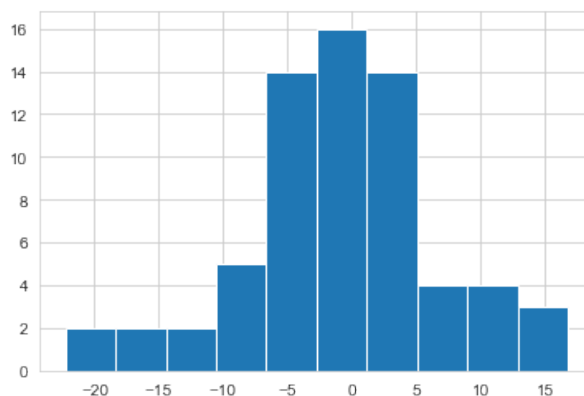


Figura 3: Diferença entre previsão e resultado verdadeiro- Percentagem de sono profundo

Tempo até o primeiro sono profundo - Previsão 2

Para a segunda previsão começou-se com o mesmo *dataset* de base em relação à previsão anterior. O processo de adição e remoção de colunas foi semelhante ao caso anterior, com apenas a exceção que desta feita foi removida a coluna **DeepSleepPercentage**, obtendo o seguinte conjunto de dados formado por apenas 4 colunas :

- **HRmean** : Média da frequência cardíaca média, meia hora antes início do sono;
- **HRmax** : Média da frequência cardíaca máxima, meia hora antes início do sono;
- **HRmin** : Média da frequência cardíaca mínima, meia hora antes início do sono;
- **DurationUntilDeepSleep** : Tempo em minutos até atingir o sono profundo;

Para realizar a previsão criaram-se diversos modelos, utilizando técnicas diferentes. Foi utilizada regressão linear, árvores de decisão, SVM, *Random Forests*, Redes Neurais simples e XGBoost.

O modelo que obteve melhor resultado foi aquele onde foi utilizada as *Random Forests*, com um MAE de 22,60 minutos. Visto a precisão da pulseira não ser muito precisa a detetar o início do sono, não parece ser um resultado mau. Na figura abaixo podemos ver as diferenças entre os valores reais e os valores previstos do modelo escolhido.

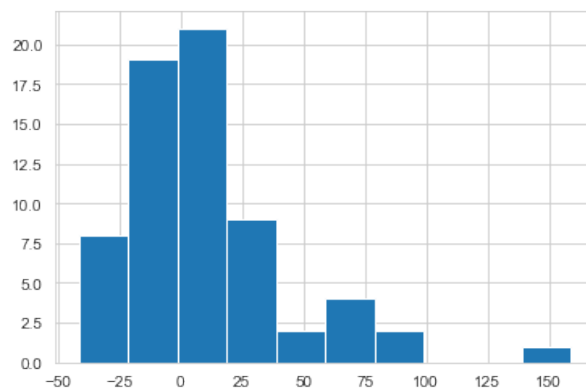


Figura 4: Diferença entre previsão e resultado verdadeiro - Minutos até primeiro sono profundo

1.6.3 Previsão de Calorias

Para finalizar tentou-se realizar uma previsão das calorias queimadas por dia, visto esta ser uma área cuja muita gente mostraria interesse.

Assim foi necessário a criação de um conjunto de dados, a partir dos dados possíveis de recolher da pulseira, que permita realizar tal previsão. As colunas do *dataset* resultante encontram-se em seguida :

- **HRmean** : Média da frequência cardíaca média ao longo do dia;

- **Hrmax** : Média da frequência cardíaca máxima ao longo do dia;
- **HRmin** : Média da frequência cardíaca mínima ao longo do dia;
- **ActiveMinutes** : Minutos ativos durante o dia;
- **Calories** : Calorias queimadas durante o dia;
- **Steps** : Número de passos realizados durante o dia;
- **Distance** : Distância percorrida ao longo do dia;
- **BeginInterval/EndInterval** : Colunas auxiliares para delimitar os dados recolhidos no prazo de um dia.

O primeiro passo foi modificar a estrutura do *dataset*, removendo todas as informações de **BeginInterval** e **EndInterval**, uma vez que a sua existência servir meramente para facilitar na criação do *dataset*.

O seguinte passo passou pela remoção de todas as linhas que tivessem algum campo vazio, visto que estes dados só influenciariam de forma negativa o modelo. Em seguida procedeu-se à remoção de *outliers* existentes nos dados numéricos.

Para realizar a previsão criaram-se diversos modelos, utilizando técnicas diferentes. Foi utilizada regressão linear, árvores de decisão, SVM, *Random Forests*, Redes Neurais e XGBoost.

O modelo que obteve melhor resultado foi aquele onde foi utilizada o SVM, com um MAE de 116 calorias. Visto que em média por dia o número total de calorias é 2000, este valor revela-se ser bastante bom. Na figura abaixo podemos ver as diferenças entre os valores reais e os valores previstos do modelo escolhido.

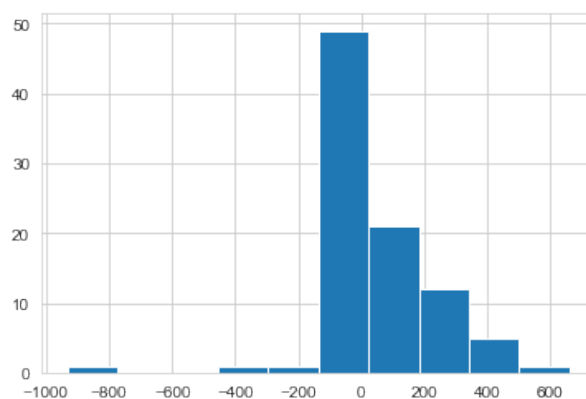


Figura 5: Diferença entre previsão e resultado verdadeiro - previsão calorias

1.7 Visualização dos Dados

Nas seguintes secções irão ser apresentadas com detalhe cada uma das páginas de visualização de dados oferecidas pela SAFit.

1.7.1 Visualização dos dados por ano

Nesta página conseguimos visualizar os dados relativos a um determinado ano. Para isso são apresentados diversos gráficos de barras com os diversos tipos de dados recolhidos com recurso à pulseira. Os gráficos são portanto os seguintes:

- Soma do número de passos dados por mês;
- Distância percorrida por mês em Quilómetros;
- Calorias queimadas por mês;
- Número de minutos ativos por mês;
- Média da pulsação média por mês;
- Média de horas dormidas por mês;
- O número de atividades físicas realizadas por mês.

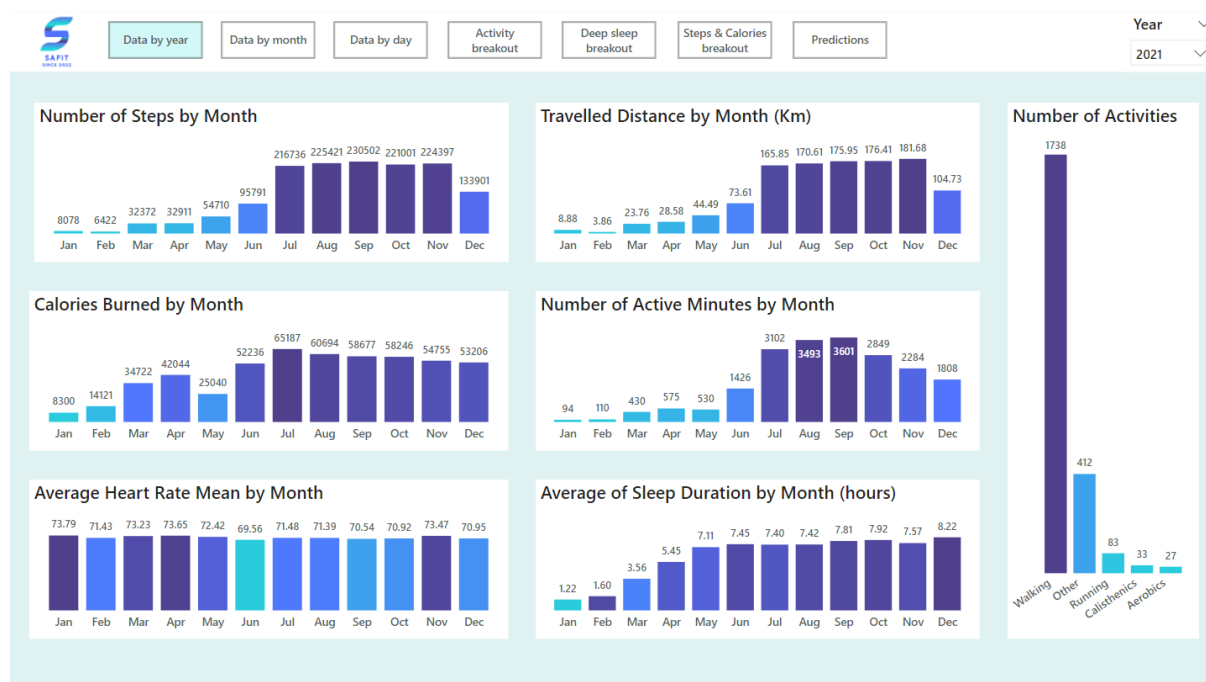


Figura 6: Dados por ano

1.7.2 Visualização dos dados por mês

Na seguinte página os dados estão desta vez organizados por mês. Desta vez não foram apenas usados gráficos de barras mas também um gráfico em *donut* e um KPI. Os gráficos representados são os seguintes:

- Número de passos realizados por dia;
- Distância percorrida por dia em kilometros;
- Número de minutos ativos por dia;
- Calorias queimadas por dia;
- Média da pulsação média no mês;
- Número de horas dormidas por dia;
- Número de atividades totais realizadas no mês.

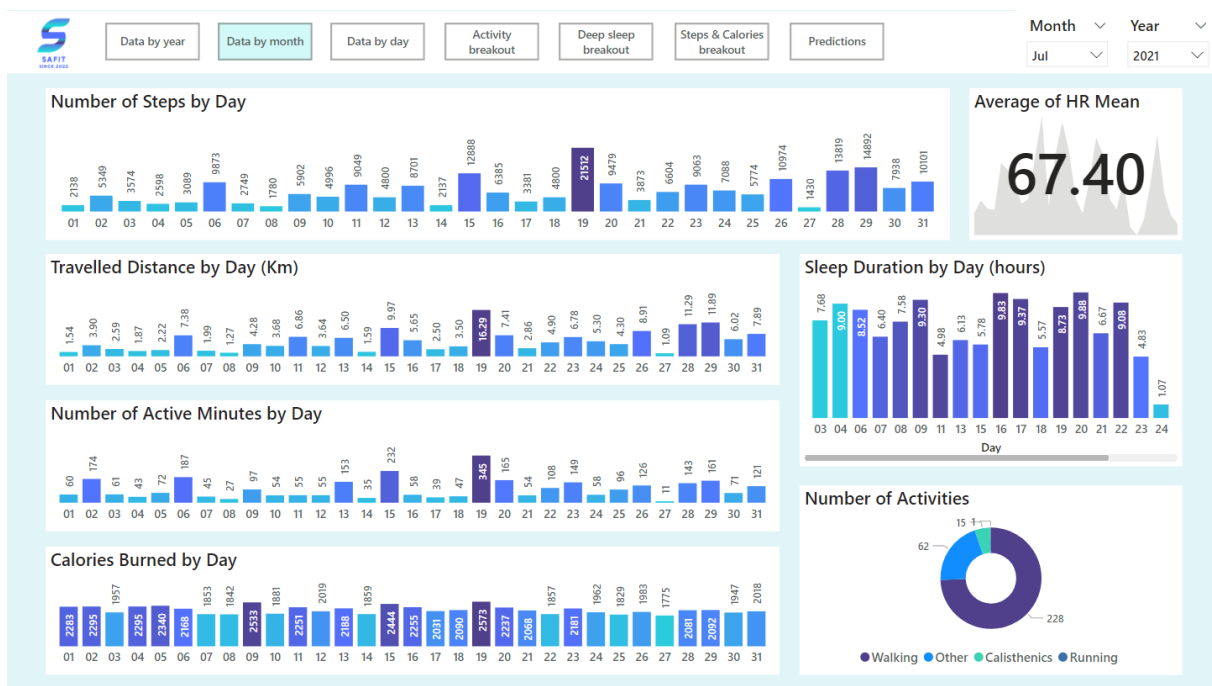


Figura 7: Dados por mês

1.7.3 Visualização dos dados por dia

Por sua vez na página ilustrada na figura 8, os dados estão demonstrados em função do dia. Para representar tais dados foram utilizados KPIs, gráficos de *donuts*, um diagrama de colunas sobrepostas e ainda gráficos de área. Os gráficos representam o seguinte:

- Número de passos realizados no dia;
- Número de calorias queimadas no dia;
- Distância percorrida em quilômetros no dia;
- Minutos ativos ao longo do dia;
- A média da pulsação média, mínima e máxima;
- O número de atividades praticadas;
- A quantidade de horas de sono profundo e sono leve;
- O número total de horas dormidas;
- A pulsação mínima ao longo do tempo;
- A pulsação média ao longo do tempo;
- A pulsação máxima ao longo do tempo.

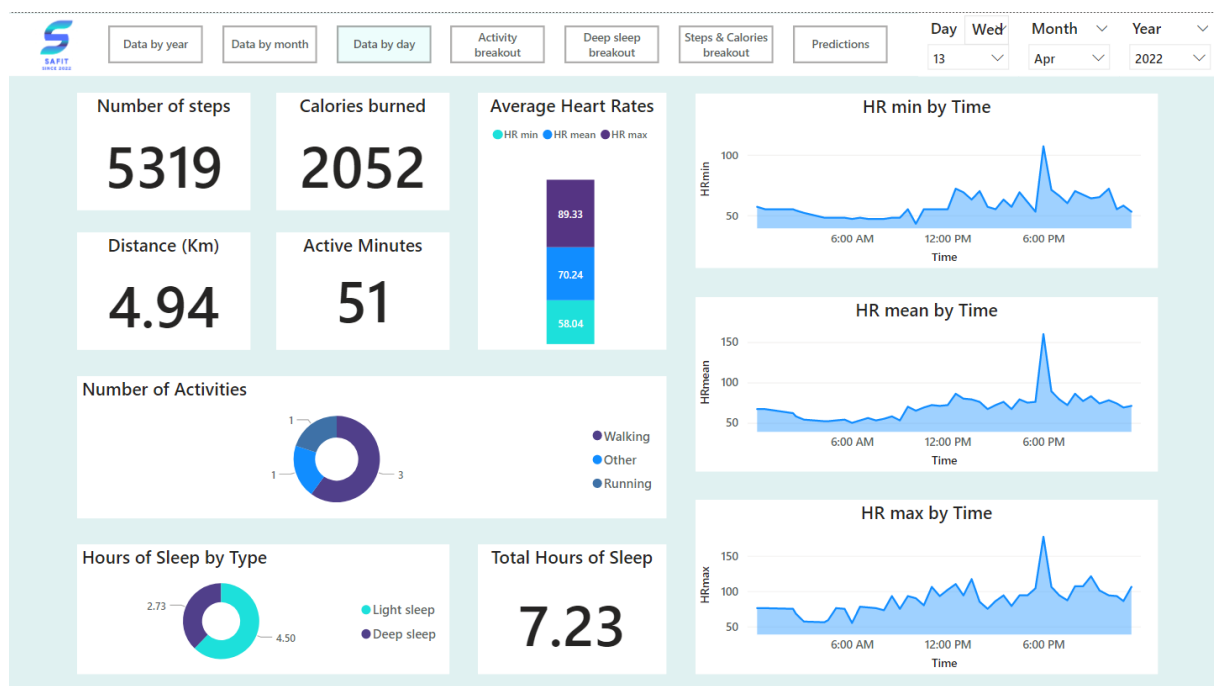


Figura 8: Dados por dia

1.7.4 Visualização detalhada das atividades realizadas

Na página em seguida é realizada uma pequena análise sobre diversos pontos aquando da realização das diversas atividades registadas pela pulseira. Assim resultaram os seguintes gráficos de barras:

- A média da pulsação mínima por atividade;
- A média da pulsação média por atividade;
- A média da pulsação máxima por atividade;
- A duração média de cada atividade;
- O número de atividades realizadas por dia da semana;



Figura 9: Dados relativos às atividades

1.7.5 Visualização detalhada do sono profundo

Em seguida estão presentes dados relativos ao sono profundo do utilizador da SAFit. Para analisar tais dados foram realizados vários gráficos de barras cujo resultado se encontra explicitado em seguida:

- Média de percentagem de sono profundo por dia da semana;
- Média de minutos necessários até atingir o sono profundo por mês;
- Média de percentagem de sono profundo por mês;
- Média de minutos necessários até atingir o sono profundo por dia da semana.

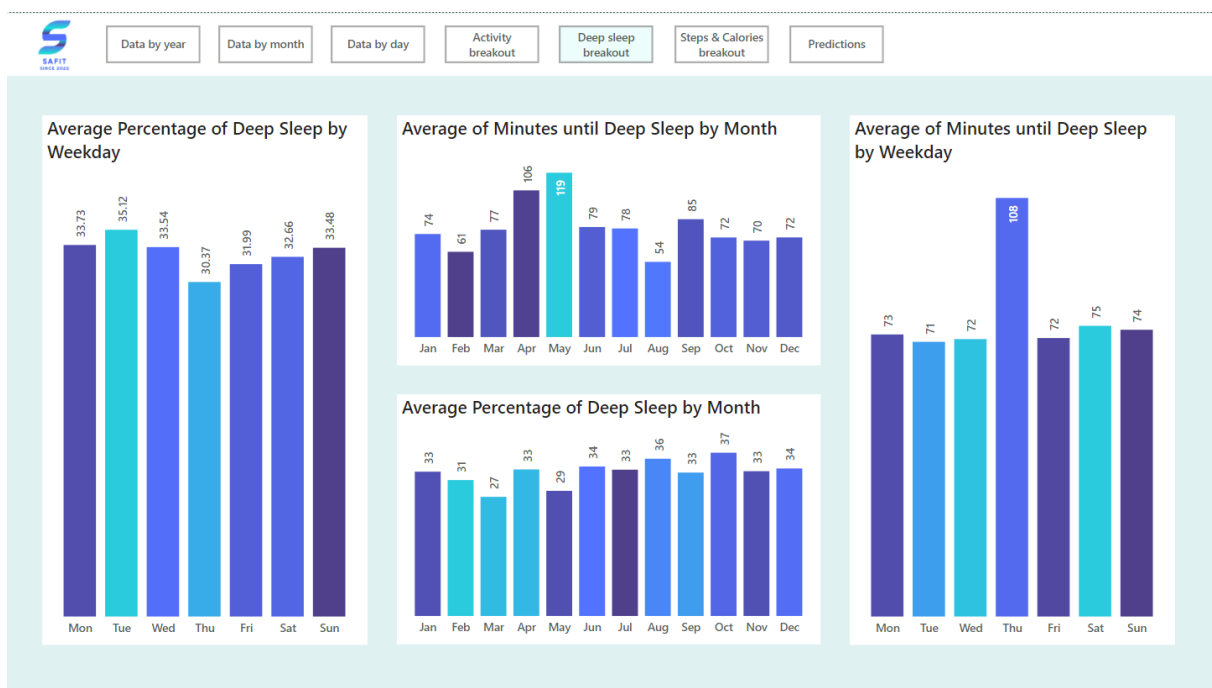


Figura 10: Dados relativos ao sono profundo

1.7.6 Visualização detalhada dos passos e calorias

Foram ainda realizados alguns testes relativamente ao número de passos e calorias gastas. Para analisar tais resultados foram utilizados os gráficos de barras e os gráficos de dispersão explicitados em seguida:

- Distância em quilómetros em função do número de passos;
- Distância em quilómetros em função das calorias queimadas;
- Minutos ativos em função de calorias queimadas;
- Média de passos por dia da semana;
- Média de calorias queimadas por dia da semana;
- Média de minutos ativos por dia da semana;



Figura 11: Dados relativos aos passos e calorias

1.7.7 Visualização das previsões

Em seguida apresenta-se a visualização dos dados das previsões dos modelos anteriormente referidos:

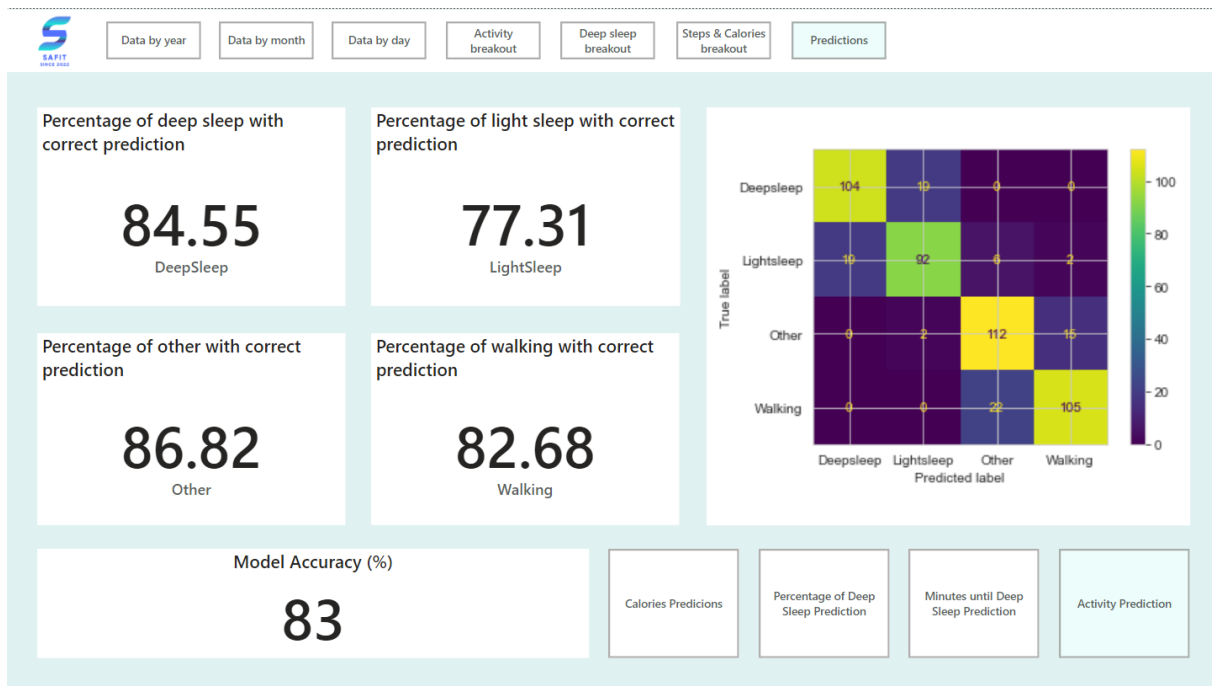


Figura 12: Visualização do modelo de previsão de atividades

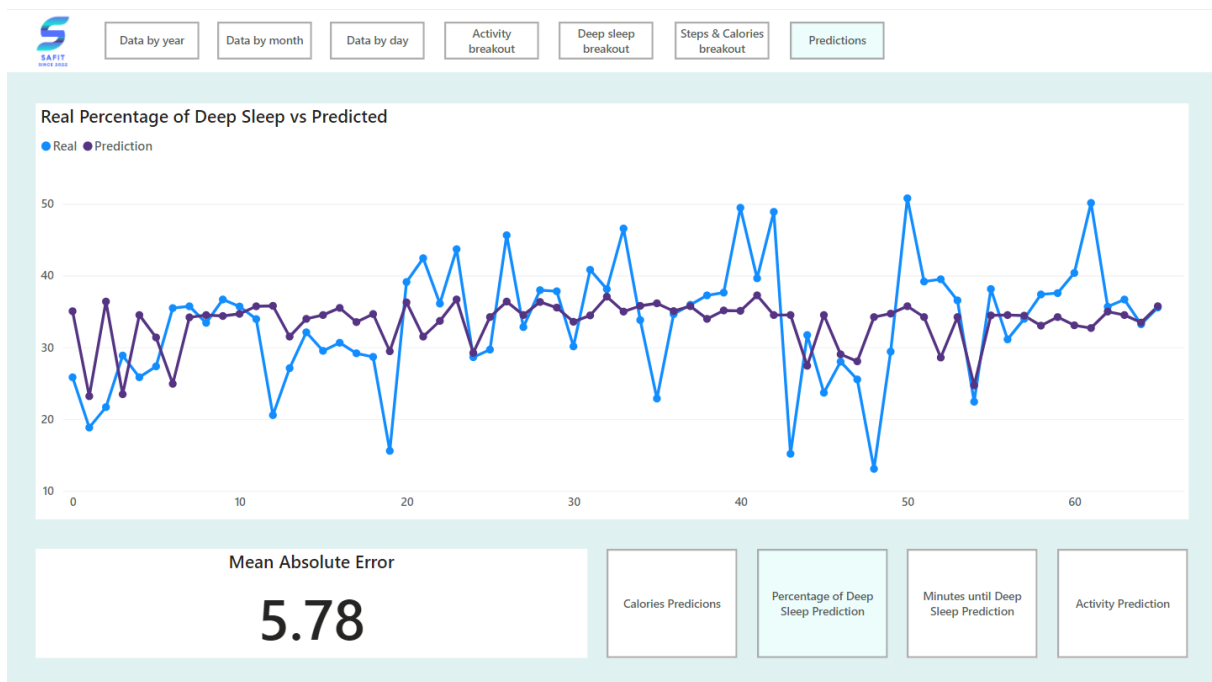


Figura 13: Visualização do modelo de previsão da percentagem de sono profundo

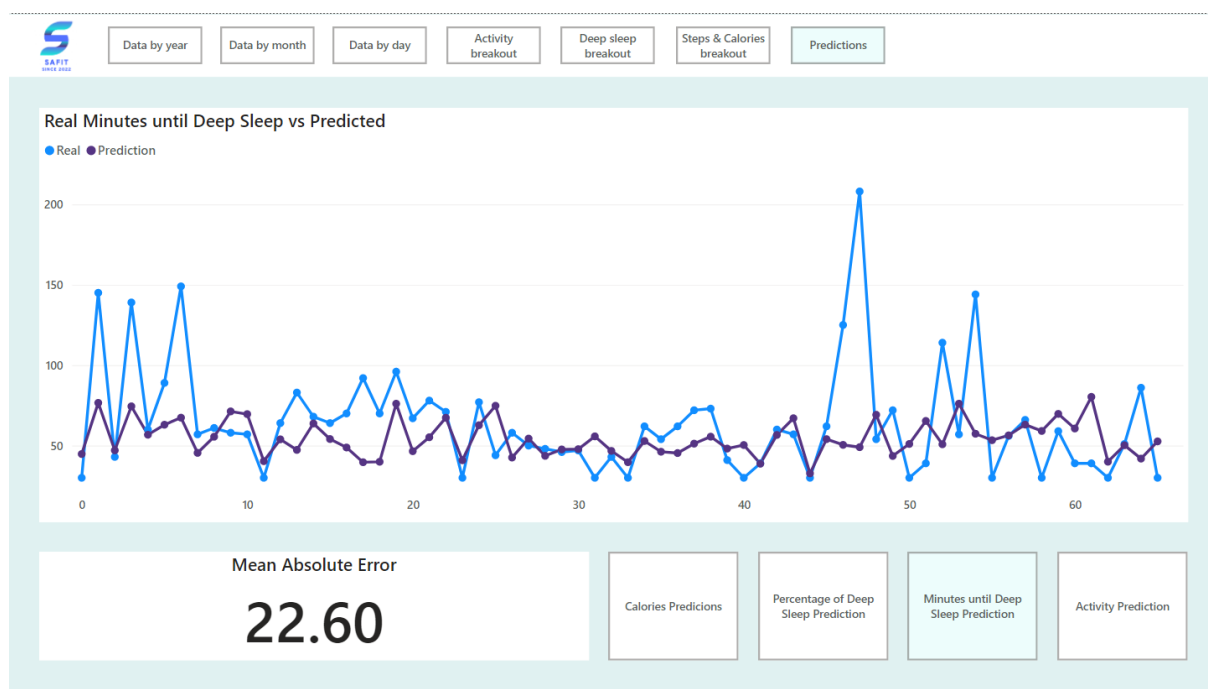


Figura 14: Visualização do modelo de previsão de minutos até atingir de sono profundo

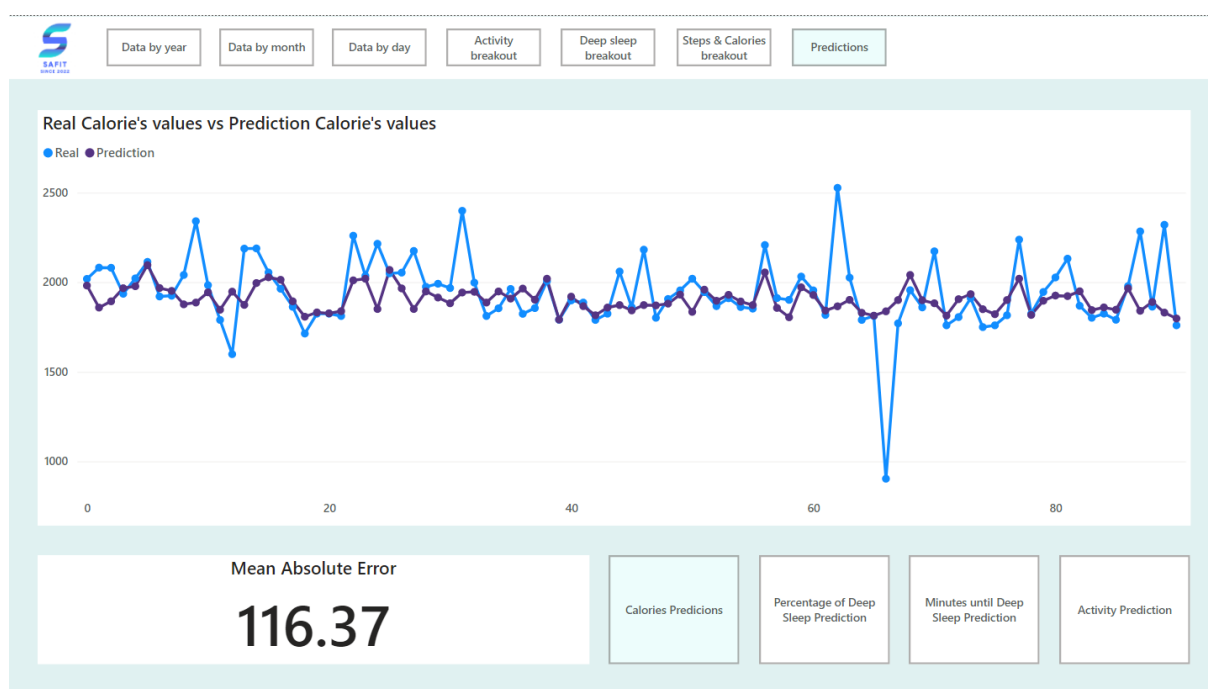


Figura 15: Visualização do modelo de previsão de calorias

1.8 Análise dos Resultados

1.8.1 Dados organizados cronologicamente

No que toca às estatísticas elementares apresentadas (número de passos, distância percorrida, número de minutos ativos, calorias queimadas, frequência cardíaca média, duração de sono e atividades) organizados por datas podemos observar que existem alguns momentos cujos dados se encontram mais incongruentes, e até um pouco estranhos, isto devido à existência de dados para recolha nesses dias ou momentos. Esta falta de dados advém do facto de o dispositivo não ter sido utilizado todos os dias, a toda a hora.

Assim sendo, os dados estão maioritariamente completos à exceção dos dados de sono dos meses de janeiro a junho de 2020, excluindo março, e em setembro do mesmo ano. A partir de junho de 2021 em diante podemos observar os resultados relevantes para todas as métricas referidas anteriormente como elementares. Nos meses aqui não mencionados, em nenhum deles foi possível obter dados completos em pelo menos 20 dos seus dias.

Focando agora nos dados mais completos, desde junho de 2021, podemos observar algumas coisas interessantes. Observando os valores dos números de passos, distância percorrida, calorias queimadas e ainda minutos ativos, conseguimos observar que o detentor da pulseira é mais ativo durante as férias do que durante o período escolar. Sendo que em 2021 o ano letivo começou em outubro, a partir dos minutos ativos, principalmente, é possível observar uma clara quebra em relação aos meses anteriores, isto muito provavelmente devido ao começo das aulas e por conseguinte aos minutos sem atividade física devido à presença nas aulas. Para além disso desde dezembro vemos uma descida abrupta, isto devido à época universitária entre os meses de dezembro e janeiro de exames e entregas.

Em fevereiro iniciou o segundo semestre na universidade, e com ele vieram menos dias de aulas para o detentor da pulseira, mais propriamente 2 ao contrário dos 3 do semestre passado que faziam o mesmo passar 4 dias na universidade em vez de apenas 2 como é o caso atual. Assim foi notável uma quebra da atividade diária já que muito do tempo passado fora de Braga por parte da fonte dos dados, excluindo fins de semana e tempo de férias apresenta uma baixa atividade já que a maior parte dos minutos ativos é devida a convívios e nesta altura as possíveis pessoas para tal, não estariam disponíveis.

No entanto em abril verifica-se um grande pico de atividade, isto mais concretamente relacionado com o aparecimento das férias da Páscoa, que não só por ser um tempo de férias o proprietário da pulseira é mais ativo, como também houve um período de viagem ao estrangeiro nesse mês, mais concretamente entre os dias 21 e 25.

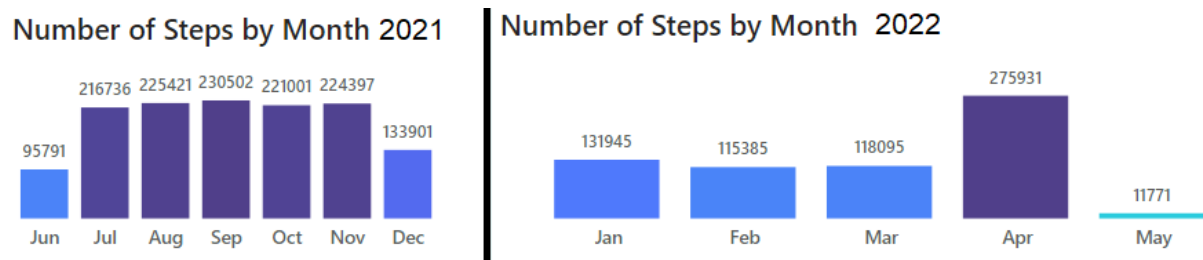


Figura 16: Comparação do número de passos

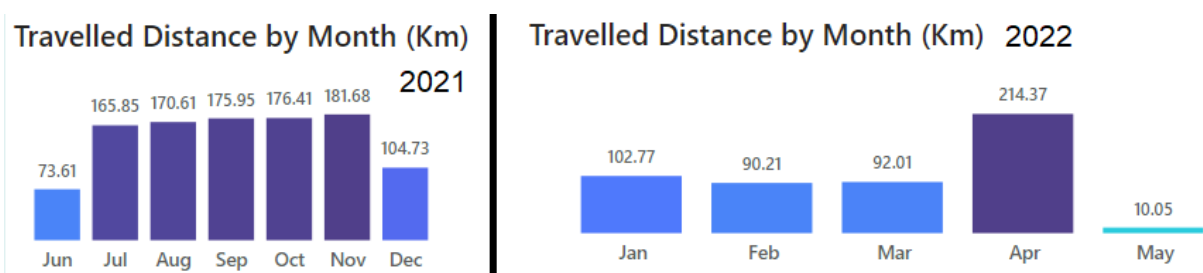


Figura 17: Comparação da distância percorrida

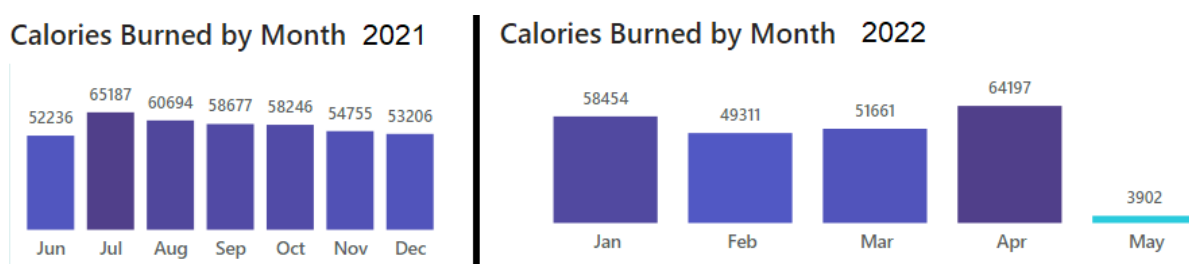


Figura 18: Comparação de calorias queimadas

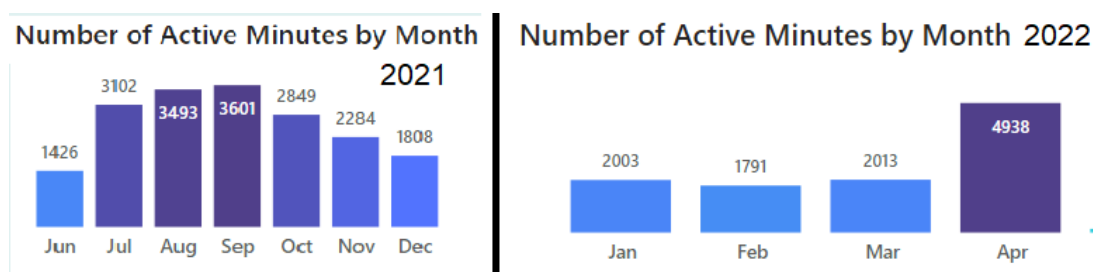


Figura 19: Comparação dos minutos ativos

No que toca a tempo total de sono os dados por mês, relativamente ao número de horas dormidas não varia muito, sempre com uma média entre valores de 7 a 8 horas aproxima-

damente, claro que existiram dias com totais de sono muito baixos, de apenas 4 horas e outras de valores mais elevados como 10 horas de sono.

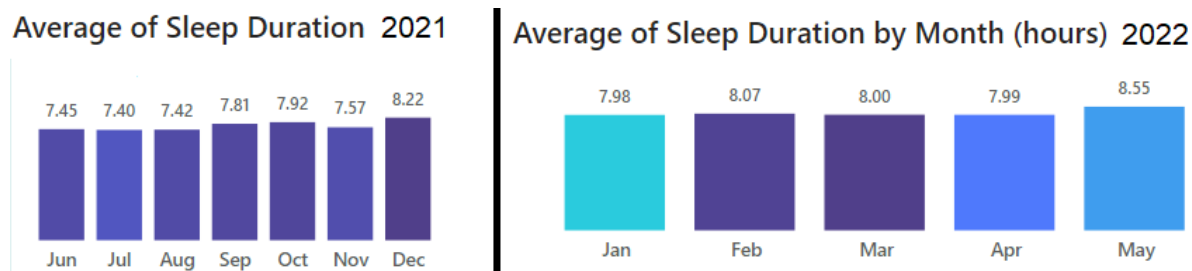


Figura 20: Comparação do tempo total de sono

Já no que toca a atividades físicas realizadas, existe sempre uma esmagadora maioria de *Walking* seguida de *Other* com alguns casos de atividades com quantidades inferiores como é o caso de *Running*, *Calisthenics* e *Aerobics*. Quanto à frequência cardíaca esta mantém-se também mais ou menos estável ao longo dos meses, oscilando entre os 70 e os 74 de frequência cardíaca média.

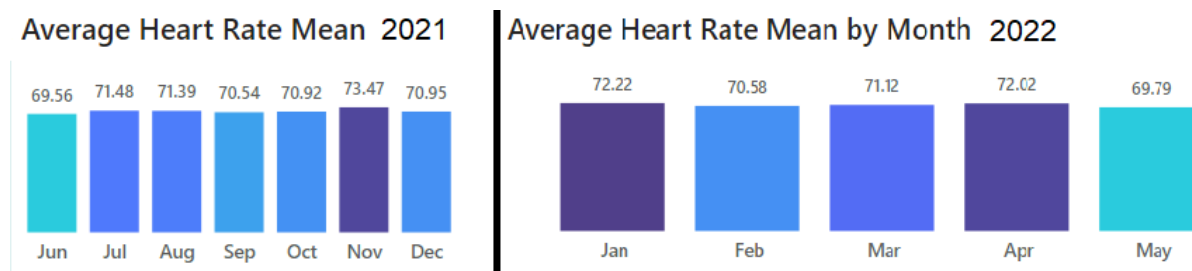


Figura 21: Comparação da frequência cardíaca

1.8.2 Dados relativos às atividades

Na figura 9 da secção da visualização de dados podemos ver diversas estatísticas sobre as atividades. Podemos ver desde logo que as atividades físicas do dono dos dados não costumam ser nem muito intensas (olhando para os gráficos das frequências cardíacas) nem costumam ser muito prolongadas, sendo que não se costumam estender mais do que 18 minutos em média.

Analisando a quantidade de atividades por dia da semana é curioso ver que segunda e terça são os dias com um maior número e os dias de fim de semana aqueles com um número inferior. Isto pode dever-se a que a maior parte das pessoas tende a tentar começar uma rotina de exercício físico no início das semanas, assim como tendem a descansar nos fins de semana.

1.8.3 Dados relativos ao sono profundo

Já sobre dados relativamente ao sono profundo na secção da visualização de dados, mais propriamente Na figura 10, podemos observar duas métricas deste tipo de sono, mais propriamente a sua percentagem do sono total e o número de minutos até atingir o primeiro destes estados durante a noite, em função do mês e do dia da semana.

Embora estas duas métricas não sejam exatamente contrárias, consegue-se observar em alguns casos uma certa correlação entre uma inferior percentagem de sono profundo e a média de minutos até chegar ao primeiro estado deste tipo.

Algo curioso são os dados relativos ao sono de quinta-feira. Esse é claramente o dia cujo sono profundo é mais afetado. Visto que o sono deste dia refere-se à noite de quarta para quinta, pode ser que estes resultados tenham um possível motivador, a existência das quartas académicas, que não favorecem com certeza o sono de uma pessoa.

1.8.4 Dados relativos aos passos e calorias

No que toca aos dados relacionados com passos e calorias da figura 11 presente na secção de visualização de dados podemos retirar algumas conclusões. Primeiro é de esperar que o número de passos e a distância percorrida sejam lineares, uma vez que uma depende quase diretamente uma da outra.

Já no que toca à comparação entre calorias queimadas com distância percorrida, assim como calorias queimadas em função de minutos ativos já não é possível observar esta linearidade, isto devido ao facto do corpo de um homem (que é o caso do dono dos dados) gastar em repouso entre 1800 e 2000 calorias em média ⁴. Deste modo, embora com bastante ruído, estas relações seguem mais uma espécie de relação logarítmica.

Já no que toca às médias por dia da semana podemos ver que as calorias queimadas não sofrem grandes diferenças, no entanto a média de passos e os minutos ativos vai variando ao longo da semana, sendo que, não exatamente da mesma maneira, isto visto que minutos ativos não necessitam obrigatoriamente de contagem de passos.

1.9 Considerações Finais

A ciência e engenharia de sensores é uma área que é relevante para praticamente todos os aspetos da vida, incluindo segurança, proteção, vigilância, monitorização, entre outros. A implementação de sensores e tecnologias de deteção tem vários benefícios, incluindo manutenção preditiva e preventiva.

Neste caso, utilizando dispositivos de monitorização de saúde e uma API conseguimos criar uma plataforma que não só mostra os padrões ou hábitos de uma pessoa, mas também tenta prever um conjunto de fatores de interesse como a forma que os batimentos cardíacos

⁴<https://blog.oficialfarma.com.br/quantas-calorias-o-corpo-queima-por-dia-com-e-sem-exercicios/>

estão relacionados com a atividade que está a ser realizada; a forma como o stress afeta a quantidade de sono profundo e o tempo até atingir este tipo de sono; e a quantidade de calorias gastas em função dos minutos ativos, passos, distância e batimentos cardíacos de um sujeito.

Tendo dito isto, os dados recolhidos foram um ponto limitador para atingir melhores resultados a nível das previsões desenvolvidas. Seja pela falta de informação derivado de um uso não contínuo do dispositivo, seja pela imprecisão dos dados, fator limitado pela qualidade e frequência de utilização dos sensores da pulseira. Além disso, para muitas das previsões, seria necessária informação do acelerómetro à qual não temos acesso, como por exemplo, saber se o utilizador se está a mover num determinado momento.

Quanto à ferramenta de visualização dos dados que utilizamos, PowerBI, também encontramos algumas dificuldades e limitações. Ao construir os gráficos notamos que por vezes os relacionamentos entre vários *datasets* não são bem estabelecidos, o que dificultou a visualização dos dados numa determinada data em específico para diversos tipos de dados em simultâneo. Em segundo lugar, tínhamos como ideia calcular previsões com base em *inputs* inseridos no PowerBI, no entanto, a ferramenta ainda não suporta a ação de correr *scripts* de Python num botão.

Com isto, podemos concluir que consideramos que houve um balanço positivo do trabalho realizado dado que as dificuldades sentidas foram superadas e foram cumpridos todos os requisitos. Foi analisado um problema, obtidos dados, desenvolvidos modelos e feita uma análise crítica adequada a esta problemática.

Como trabalho futuro, pretendia-se melhorar a diversidade dos dados e a qualidade dos modelos elaborados.