

UNIVERSIDADE DO MINHO
MESTRADO INTEGRADO EM ENGENHARIA INFORMÁTICA

SISTEMAS AUTÓNOMOS

SISTEMAS INTELIGENTES

(2º SEMESTRE / 4º ANO)

Monitorização de parâmetros ambientais para a prática de desporto

GRUPO 10

Diogo Braga (a82547)
Pedro Ferreira (a81135)
Ricardo Caçador (a81064)
Ricardo Veloso (a81919)

Março, 2020

Conteúdo

| | | |
|----------|--|-----------|
| 1 | Introdução | 2 |
| 2 | Fundamentação teórica | 3 |
| 3 | Sensorização e armazenamento dos dados | 4 |
| 3.1 | Open Weather Map | 4 |
| 3.2 | Open Air Quality | 6 |
| 3.3 | Open UV | 7 |
| 3.4 | Firestore | 8 |
| 4 | Automatização do processo de recolha de dados | 9 |
| 5 | Processamento dos dados | 11 |
| 5.1 | Estruturação dos dados | 11 |
| 5.2 | Normalização dos dados para previsão | 12 |
| 6 | Modelos de previsão | 13 |
| 6.1 | Avaliação da eficiência do ajuste | 13 |
| 6.2 | Previsão <i>vs</i> Realidade | 15 |
| 6.3 | Previsão - 24 horas | 18 |
| 6.4 | Classificação e armazenamento | 19 |
| 7 | Interface do sistema | 20 |
| 7.1 | Dados sensorizados do clima | 20 |
| 7.2 | Dados sensorizados da radiação ultravioleta | 22 |
| 7.3 | Previsão das melhores horas para prática de desporto | 23 |
| 8 | Conclusão | 24 |
| | Referências | 25 |

1 Introdução

O presente relatório detalha a criação de um sistema relacionado com Inteligência Ambiente e Sensorização. Este foi proposto no âmbito da Unidade Curricular de Sistemas Autónomos, inserida no perfil de especialização em Sistemas Inteligentes.

Diariamente, um grande número de cidadãos sai à rua com intenções de praticar desporto. Com efeito, são cada vez mais os espaços públicos existentes que se destinam à realização de atividade física. Tendo em vista a promoção de um estilo de vida mais saudável, é totalmente aconselhado que este seja praticado em condições favoráveis e, principalmente, que não coloque a saúde dos praticantes em risco. No entanto, existem fatores que condicionam e desaconselham a prática de atividade física, como, por exemplo: elevadas temperaturas - que podem levar a drásticas desidratações do corpo humano; a fraca qualidade do ar - que pode prejudicar o sistema respiratório; e, ainda, a elevada radiação ultravioleta - que pode provocar o desenvolvimento de graves doenças de foro dermatológico.

Procurando informar a população sobre quais os momentos ideais para a prática de desporto, surgiu o sistema apresentado neste relatório. O principal objetivo do sistema é oferecer aos utilizadores dados sobre quais os melhores momentos numa semana para praticar desporto, tendo em conta as condições ideais de variados parâmetros ambientais.

Além da motivação apresentada, é também factual o aumento abrupto das taxas de sedentarismo e obesidade. Recorrentemente, são lançadas notícias e artigos de inúmeras áreas ligadas à saúde e ao bem-estar físico que aconselham e motivam a prática de desporto. Um dos principais entraves a uma maior taxa de atividade física, principalmente no grupo da população sedentária e daqueles que raramente praticam desporto, são eventuais desconfortos sentidos durante o exercício físico. Entre outros fatores, estes podem ser provocados, por exemplo, pelo calor excessivo. Assim, este sistema surge também na lógica de combater estas situações, procurando indicar ao seu utilizador um período ótimo para exercitar. Certamente, quanto mais propícias foram as condições ambientais envolventes, melhor será qualidade da experiência. Desta forma, poderemos ajudar a disseminar a prática de atividade desportiva, ultrapassando os entraves supramencionados.

Em suma, o sistema tem como objetivo aconselhar e apresentar ao utilizador quais os momentos em que as condições ambientais são favoráveis à prática desportiva, tendo por base os níveis da temperatura, da humidade, da qualidade do ar, e também da radiação ultra-violeta.

2 Fundamentação teórica

Todos aqueles que praticam exercício físico com regularidade sabem que as condições meteorológicas podem afetar a performance do praticante. Apesar de ser possível praticar desporto nas mais diferentes condições meteorológicas, até mesmo nas mais extremas, através dos vários estudos já feitos é possível perceber que existem certas condições, quer a nível da qualidade do ar como em parâmetros mais gerais como temperatura e humidade, que permitem aos atletas praticar desporto de uma maneira mais eficiente e saudável.

O fator mais influenciador, é sem dúvida a temperatura. De acordo com um estudo efectuado em 2010 sobre o impacto das condições do ambiente na performance dos maratonistas, chegou-se à conclusão que as temperaturas ideais encontram-se entre os 10 e os 20 graus Celsius. As temperaturas entre os 0 e os 10 graus Celsius e os 20 e 26 graus Celsius foram consideradas aceitáveis. Todas as outras temperaturas de extremos revelaram-se prejudiciais para a saúde dos atletas [1].

O segundo factor que mais influencia a performance dos atletas é a humidade. Certos valores de humidade no ar combinados com a temperatura podem influenciar negativamente a saúde de um atleta. Para perceber essa relação foram feitos vários estudos analisando o volume máximo de oxigénio que o corpo consome durante o exercício físico, mais conhecido no meio como VO2 máximo. Para ter uma noção dos valores de humidade considera-se um tempo seco quando a humidade está a cerca de 30% enquanto que com os valores da humidade à volta dos 60% já estamos perante um tempo considerado húmido. Em temperaturas elevadas e húmidas o VO2 máximo de um atleta pode descer mais de 12%. Já se a humidade for um pouco mais baixa mas temperatura for semelhante ocorrem descidas no VO2 máximo de cerca de 5%. Quanto maior a humidade mais dificuldade o atleta vai ter em respirar e maior vai ser o seu ritmo cardíaco. Quanto mais húmido o ar estiver, mais água existe no ar e por consequência, mais complicada é a evaporação do suor não permitindo ao corpo do atleta refrescar. Assim sendo, para valores de humidade entre os 60 e os 90% uma corrida de ritmo médio de 8 minutos vai parecer uma corrida de 6 minutos pois o coração vai bater mais forte e mais rapidamente devido à humidade. Concluindo, consideram-se os valores de humidade ideais abaixo dos 50% [2].

A qualidade do ar, não tendo um grande impacto a curto prazo, é importante pois pode ser muito prejudicial a médio-longo prazo para a saúde do desportista. O ozono, o dióxido de nitrogénio e o dióxido de enxofre são os 3 parâmetros cujos valores iremos analisar. Destes 3 valores apenas o ozono e o dióxido de nitrogénio têm impacto direto na performance. É difícil perceber o impacto direto de apenas um poluente pois nunca está isolado. As concentrações de ozono aumentam em proporção linear com a temperatura ambiente confirmando assim que em temperaturas mais baixas a saúde do atleta é favorecida pois não há grande concentração de ozono. Segundo o estudo no qual nos baseamos os níveis destes 3 parâmetros apresentados encontravam-se em níveis normais apesar de serem maratonas em grandes cidades pelo que chegou-se à conclusão que não existe um impacto alto [3] [4] [5].

Relativamente à radiação ultravioleta, todos sabemos os danos que esta pode causar à nossa pele. Realizamos então uma pesquisa mais focada nos valores que esta pode tomar e os perigos que estes valores representam. Em primeiro lugar, um índice UV superior a 5.9 é considerado prejudicial para a saúde logo, independentemente de qualquer outro fator, devemos evitar praticar exercício físico nestas circunstâncias. Em segundo lugar, um índice UV entre 3.0 e 5.9 representa um risco moderado logo, dependendo dos valores dos fatores anteriormente mencionados, poderemos ou não praticar exercício físico. Finalmente, um índice UV menor que 3.0 não proíbe qualquer tipo de atividade física no exterior já que o perigo que lhe é atribuído é baixo [6]

3 Sensorização e armazenamento dos dados

Tendo em conta as necessidades apresentadas pelo sistema, concluiu-se ser necessário realizar sensorização ao nível das condições climáticas já apresentadas: temperatura, humidade, qualidade do ar e radiação ultra-violeta.

Devido à falta de sensores físicos, este processo é efetuado de forma virtual, sendo necessário extrair dados de APIs que disponibilizem essa informação. Nesse sentido, são acedidas as seguintes APIs:

- Open Weather Map;
- Open AQ (Air Quality);
- Open UV (Ultraviolet).

Para este projeto em específico, a sensorização incidirá sobre cidade de Braga. Assim, todos os dados referentes às condições climáticas dizem respeito a cidade em específico. No **Open Weather Map** foi utilizado o **id** da cidade atribuída pelo próprio website (2742032), sendo que, no **Open AQ** foi utilizado o mesmo método (PT01042). Diferenciadamente, mas com a mesma intenção, no **Open UV** foram utilizadas as coordenadas do centro da cidade, sendo estas as seguintes: (41.55032, -8.42005).

A informação disponibilizada pelas APIs possui um papel fulcral na concretização das funcionalidades do sistema. Deste modo, são realizados **http requests** a cada um dos *websites* associados às APIs, passando, através do **url**, a identificação da cidade como parâmetro.

Uma vez recebida a resposta da API, procede-se a filtrar a informação obtida, recolhendo apenas a necessária para fornecer dados ao sistema. Assim, é realizada uma primeira fase de processamento dos dados: é descartada a informação que não é útil ao sistema, sendo transmitidos ao back-end apenas os dados estritamente necessários.

De forma a construir uma base de registos sólida, sobre a qual seja possível efetuar previsões aplicando modelos de *machine learning*, as APIs enunciadas são consultadas de hora em hora, sendo os resultados guardados na **Cloud Firestore**. Esta base de dados encontra-se organizada por coleções que contêm documentos com campos no seu interior. Para maior modularidade e facilidade de leitura dos dados, cada coleção é referente a uma API. Por sua vez, cada documento diz respeito aos dados extraídos da resposta a um pedido, sendo a sua chave identificadora a hora a que o pedido foi realizado.

De seguida, nas secções 3.1 a 3.3 são apresentados exemplos de respostas obtidas de cada uma das APIs. Como foi referido, apenas alguns campos da resposta são persistidos e mapeados para a base de dados. Assim sendo, apresenta-se também a estrutura dos documentos que compõem as diferentes coleções. Na secção 3.4 é apresentada uma visão geral sobre a base de dados.

3.1 Open Weather Map

Para efeitos de demonstração, é apresentado, de seguida, o resultado obtido de um pedido **http** à API do **Open Weather Map**, incidente sobre a cidade de Braga:

```
1 {  'base': 'stations',
2    'clouds': {'all': 0},
3    'cod': 200,
4    'coord': {'lat': 41.55, 'lon': -8.42},
5    'dt': 1585311929,
6    'id': 2742032,
7    'main': {  'feels_like': 283.45,
8               'humidity': 44,
9               'pressure': 1011,
10              'temp': 285.93,
11              'temp_max': 285.93,
12              'temp_min': 285.93},
13    'name': 'Braga',
14    'sys': {  'country': 'PT',
15             'id': 57094,
16             'sunrise': 1585290269,
17             'sunset': 1585335190,
18             'type': 3},
19    'timezone': 0,
20    'weather': [ {  'description': 'clear sky',
21                   'icon': '01d',
22                   'id': 800,
23                   'main': 'Clear'}],
24    'wind': {'deg': 56, 'gust': 5.81, 'speed': 0.89}}
```

Listing 1: Exemplo de um *http request* realizado ao Open Weather Map, para a cidade de Braga

A coleção criada a partir desta informação é a 'WM', sendo os documentos constituídos pelos seguintes atributos:

- **feels_like:** parâmetro de temperatura responsável pela percepção humana do clima, em Kelvin;
- **general_weather:** condições climáticas;
- **humidity:** humidade, em percentagem;
- **pressure:** pressão atmosférica (ao nível do mar), em hPa;
- **temp:** temperatura efetiva no momento, em Kelvin;
- **temp_max:** temperatura máxima no momento (desvio da temperatura atual que é possível para grandes cidades e megalópoles expandidas geograficamente), em Kelvin;
- **temp_min:** temperatura mínima no momento (desvio da temperatura atual que é possível para grandes cidades e megalópoles expandidas geograficamente), em Kelvin;

- **wind_speed**: velocidade do vento, em metro/s.

3.2 Open Air Quality

De seguida, apresenta-se o resultado de um pedido `http` à API do Open AQ:

```
1 {  'meta': {  'found': 1,
2          'license': 'CC BY 4.0',
3          'limit': 100,
4          'name': 'openaq-api',
5          'page': 1,
6          'website': 'https://docs.openaq.org/'},
7  'results': [ {  'cities': ['Braga'],
8                'city': 'Braga',
9                'coordinates': {  'latitude': 41.56944,
10                               'longitude': -8.45694},
11                'count': 40121,
12                'country': 'PT',
13                'countsByMeasurement': [ {  'count': 12036,
14                                           'parameter': 'no2'},
15                                           {  'count': 14363,
16                                           'parameter': 'o3'},
17                                           {  'count': 13722,
18                                           'parameter': 'pm10'}],
19                'firstUpdated': '2017-09-23T15:00:00.000Z',
20                'id': 'PT-48',
21                'lastUpdated': '2020-04-03T06:00:00.000Z',
22                'location': 'PT01042',
23                'locations': ['PT01042'],
24                'parameters': ['no2', 'o3', 'pm10'],
25                'sourceName': 'EEA Portugal',
26                'sourceNames': ['EEA Portugal'],
27                'sourceType': 'government',
28                'sourceTypes': ['government']}]}
```

Listing 2: Exemplo de um *http request* realizado ao Open AQ, para a cidade de Braga

Analisando o resultado, facilmente se conclui que a informação relevante a ser extraída corresponde ao valor dos parâmetros `no2`, `o3` e `pm10`. Assim, é criada a coleção 'AQ', onde cada documento apresenta os atributos:

- **no2**: medição do dióxido de nitrogénio/azoto;

3.3 Open UV

- **o3**: medição do ozônio/ozono;
- **pm10**: medição das partículas inaláveis, de diâmetro inferior a 10 micrómetros.

3.3 Open UV

Em relação à API do Open UV, a resposta recebida é a seguinte:

```
1 {  'ozone': 389.2,
2    'ozone_time': '2020-04-03T09:07:25.360Z',
3    'safe_exposure_time': {  'st1': 46,
4                             'st2': 56,
5                             'st3': 74,
6                             'st4': 93,
7                             'st5': 149,
8                             'st6': 279},
9    'sun_info': {  'sun_position': {  'altitude': 0.7707166232697027,
10                                     'azimuth': -0.8175392231459628},
11                 'sun_times': {  'dawn': '2020-04-03T05:45:54.600Z',
12                                 'dusk': '2020-04-03T19:30:03.027Z',
13                                 'goldenHour': '2020-04-03T18:25:16.932Z',
14                                 'goldenHourEnd': '2020-04-03T06:50:40.696Z',
15                                 'nadir': '2020-04-03T00:37:58.814Z',
16                                 'nauticalDawn': '2020-04-03T05:12:35.381Z',
17                                 'nauticalDusk': '2020-04-03T20:03:22.246Z',
18                                 'night': '2020-04-03T20:37:54.750Z',
19                                 'nightEnd': '2020-04-03T04:38:02.877Z',
20                                 'solarNoon': '2020-04-03T12:37:58.814Z',
21                                 'sunrise': '2020-04-03T06:13:59.881Z',
22                                 'sunriseEnd': '2020-04-03T06:16:52.420Z',
23                                 'sunset': '2020-04-03T19:01:57.747Z',
24                                 'sunsetStart': '2020-04-03T18:59:05.208Z'}}},
25    'uv': 3.5846,
26    'uv_max': 5.0823,
27    'uv_max_time': '2020-04-03T12:37:58.814Z',
28    'uv_time': '2020-04-03T10:30:37.143Z'}
```

Listing 3: Exemplo de um *http request* realizado ao Open UV, para a cidade de Braga

A coleção criada a partir desta informação é a 'UV', cujos documentos apresentam os seguintes atributos:

- **uv**: índice UV (tempo real);

- **uv_time:** data e hora do índice UV em UTC;
- **uv_max:** índice UV máximo para o dia (ao meio-dia solar);
- **uv_max_time:** data e hora máxima do índice UV (meio-dia solar), em UTC;
- **st1:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 1 (*Fitzpatrick Skin Types*);
- **st2:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 2 (*FST*);
- **st3:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 3 (*FST*);
- **st4:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 4 (*FST*);
- **st5:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 5 (*FST*);
- **st6:** tempo de exposição segura (minutos até à queimadura) para o tipo de pele 6 (*FST*).

3.4 Firestore

De seguida, apresenta-se o panorama geral da *dashboard* de gestão da base de dados, no Cloud Firestore. Através desta, é possível visualizar as coleções armazenadas, assim como os diversos documentos que as compõem, sendo ainda possível visualizar a organização e os dados destes. Foram vantagens como estas que nos levaram a integrar este software no back-end da aplicação em construção.

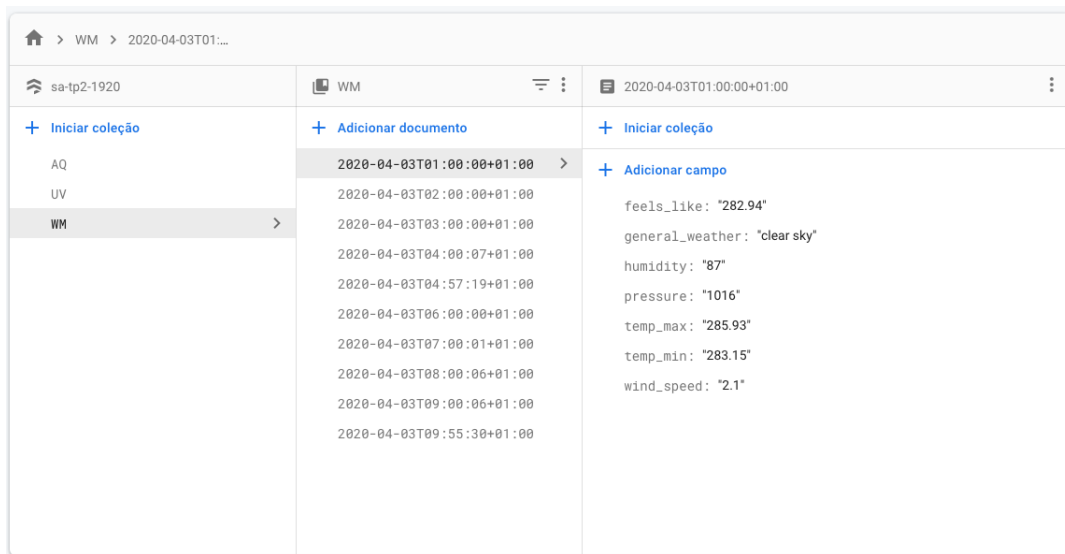


Figura 1: Vista geral da interface do Firestore

Na figura, em particular, visualizam-se as três coleções estabelecidas como necessárias para a conceção da aplicação ('AQ': Open AQ; 'UV': Open UV; 'WM': Open Weather Map). Além das coleções, visualiza-se a utilização das datas como chaves associadas a cada documento e, em cada um destes documentos, os campos com os atributos que foram mapeados das respetivas respostas das APIs.


4 Automatização do processo de recolha de dados

Uma vez que era pretendido que fosse gerado um conjunto significativo de dados, deparamo-nos com duas opções: realizar pedidos de dados históricos, isto é, construir um *script* que, por exemplo, realizasse um pedido referente a todas as horas das últimas 3 semanas; ou, automatizar o processo de recolha de dados, executando, em intervalos de tempo regular, um *script* responsável por obter dados referentes àquele instante. Como já foi referido, é feita uma chamada às APIs a cada hora, pelo que seguimos a segunda opção. Esta escolha foi motivada pelo facto de ser a que mais se aproxima do comportamento de um sensor físico. Desta forma, podemos abstrair o processo de obtenção de dados como sendo um conjunto de sensores que, a cada hora, enviam registos para uma base de dados na *cloud*.

Seguindo esta abordagem, surgiu o problema de como executar o *script* em intervalos regulares de tempo. Uma solução simples seria a utilização de *cron jobs* (comando Unix para o agendamento de tarefas). No entanto, seria necessário manter um servidor ligado permanentemente, de forma a correr o *script* no horário definido. Tal seria impensável utilizando os nossos computadores pessoais, pelo que ponderamos alocar uma máquina virtual no *Compute Engine* da *Google Cloud Platform* (GCP). Usufruindo do *plafond* destinado ao período experimental de utilização desta plataforma, esta solução seria economicamente viável. Contudo, não se justificava o custo de manter uma instância ligada permanentemente, com um tempo de computação tão reduzido (cerca de 10ms por hora).

Procurando uma melhor solução para o problema, decidimos optar por instalar o *script* de obtenção de dados numa *Cloud Function*. Deste processo resultou um *endpoint* que, ao receber um pedido **http GET**, executa o *script* desenvolvido para obtenção dos dados. Este pode ser acedido através do seguinte URL: https://europe-west3-kinetic-silicon-273019.cloudfunctions.net/API_Fetch. Posto isto, era apenas necessário enviar o pedido **http** a cada hora. Para implementar este comportamento, usufruindo das ferramentas disponibilizadas pela GCP, recorremos ao *Cloud Scheduler*, um SaaS para *cron jobs*.

Na seguinte figura, apresenta-se o *job* implementado no *Cloud Scheduler* para realizar o pedido à *Cloud Function* desenvolvida:





Cloud Scheduler


Jobs


[+ CREATE JOB](#)


[REFRESH](#)

 EDIT

 PAUSE

 RESUME

 DELETE



Filter jobs

| <input type="checkbox"/> | Name ↑ | State | Description | Frequency | Target | Last run |
|--------------------------|-----------|---------|-------------|------------------------------|---|-------------------------|
| <input type="checkbox"/> | API_Fetch | Enabled | | 0 * * * * (Europe/Lisbon) | URL : https://europe-west3-kinetic-silicon-273019.cloudfunctions.net/API_Fetch | 3 Apr 2020, 12:00:00 |

Figura 2: *Job* para envio de pedido **http GET** à *Cloud Function*

Com a conjugação destas duas ferramentas, conseguimos solucionar o problema de execução do *script* em intervalos de tempo regulares. Além disso, estamos a minimizar os gastos associados, reduzindo ao máximo o tempo de computação associado a este processo. Com efeito, na seguinte figura podemos constatar essa situação:



Figura 3: Padrão de invocação da *Cloud Function*

Os traços a azul representam as chamadas à função. Assim, podemos concluir que os dados estão a ser recolhidos da forma pretendida.

5 Processamento dos dados

Nesta secção vão ser enunciadas as técnicas de processamento realizadas sobre os dados, como a extração destes da base de dados e a atribuição de tipos. Esta fase apresenta-se de extrema importância visto que possui total influência no desempenho dos modelos de previsão criados na fase seguinte do projeto.

Tendo em conta a fundamentação apresentada na secção 2, os fatores selecionados como principais a analisar de forma a encontrar as melhores condições para a prática de desporto foram os seguintes:

- Feels_like;
- Humidity;
- Uv.

Esta capítulo assenta, portanto, sobre a estruturação dos dados e sobre a normalização das principais *features*, estas que vão ser as utilizadas nos modelos de previsão.

5.1 Estruturação dos dados

Numa fase inicial, ao obter os dados do back-end, é possível constatar que estes se encontram estruturados do mesmo método que se encontravam quando extraídos do `http request`. Para obter um bom aproveitamento dos dados é necessário que estes possuam qualidade, no entanto, da forma como estes se encontram no momento da extração, tal aproveitamento não é possível alcançar.

Posto isto, para conseguir adaptar os dados extraídos ao nosso modelo de aprendizagem, foi criado um objeto `DataFrame`, já com os `datatypes` corretamente atribuídos, para os dados do `OpenWeatherMap` e para os dados do `Open UV`.

- Da coleção 'WM' foram aproveitadas e estabelecidas as seguintes colunas e respetivos objetos:
 - **feels_like**: float64 obtido do parâmetro 'feels_like' (subtração do valor 273.15);
 - **general_weather**: object (sem alteração);
 - **humidity**: int64 obtido do parâmetro 'humidity';
 - **pressure**: int64 obtido do parâmetro 'pressure';
 - **temp**: float64 obtido do parâmetro 'temp' (subtração do valor 273.15);
 - **temp_min**: float64 obtido do parâmetro 'temp_min' (subtração do valor 273.15);
 - **temp_max**: float64 obtido do parâmetro 'temp_max' (subtração do valor 273.15);
 - **wind_speed**: float64 obtido do parâmetro 'wind_speed';
 - **wind_degree**: float64 obtido do parâmetro 'wind_degree';
- Da coleção 'UV' foram aproveitadas e estabelecidas as seguintes colunas e respetivos objetos:
 - **uv**: float64 obtido do parâmetro 'uv';
 - **uv_time**: object (sem alteração);
 - **uv_max**: float64 obtido do parâmetro 'uv_max';
 - **uv_max_time**: object (sem alteração);
 - **st1**: object (sem alteração);

- **st2:** object (sem alteração);
- **st3:** object (sem alteração);
- **st4:** object (sem alteração);
- **st5:** object (sem alteração);
- **st6:** object (sem alteração);

5.2 Normalização dos dados para previsão

Com base nas *features* selecionadas como principais, foi criado um dataframe para cada uma contendo todos os valores armazenados desde o primeiro momento de sensorização. Esta informação apresenta-se como fulcral tendo em conta que são estes os dados históricos que vão proporcionar a aprendizagem dos modelos de previsão.

De forma a proporcionar um melhor desempenho aos modelos criados para a previsão, os dados passaram por um processo de normalização. Para tal, recorremo ao `MinMaxScaler` da biblioteca `sklearn`. Esta normalização dimensiona e transforma cada *feature* individualmente, escalando cada uma para um determinado intervalo e facilitando, assim, o processo de aprendizagem dos modelos de *machine learning* aplicados.

6 Modelos de previsão

Como forma de realizar as previsões, e tendo em conta que a aprendizagem é um processo essencial neste projeto, foram aplicados como modelos as Redes Neurais Artificiais. As *Multilayer Perceptron (MLP)*, apesar de muito utilizadas neste ramo, não se apresentaram como viáveis visto não possuírem noção de sequência e tempo. Estas características são dadas como cruciais para realizar previsão de uma série temporal, e neste sentido, foram procuradas outros tipos de rede.

Desta pesquisa surgiu a decisão de aplicar *Long Short-Term Memory Networks (LSTMs)*. Nestas redes o conhecimento é transferido de iteração em iteração, existindo por isso memória e uma noção do tempo. Este tipo de rede foi, portanto, o escolhido para aplicar no problema apresentado neste projeto. Uma vez que esta foi a primeira que trabalhamos com este tipo de redes, optamos por criar 3 redes, todas com a mesma arquitetura, isolando assim a previsão de cada uma das *features* que são utilizadas para determinar quais os melhores momentos para a prática desportiva. Desta forma, conseguimos reduzir a complexidade do problema.

Para efeitos de treino, foi necessário modificar as *dataframes* mencionadas no capítulo anterior. Uma vez que estamos perante um problema de previsão de uma série temporal, tivemos que adaptar o *dataset* dos dados recolhidos pelo sensores a esta temática. Assim, transformamos as *dataframes* em conjuntos de 169 registos, (168 registos correspondem às 168 horas da última semana, sendo que se procura prever o valor da 169^a hora).

A mesma lógica é aplicada para efeitos de previsão das próximas 24 horas. São utilizados como *input* os valores referentes às 168 horas anteriores ao momento em que se executa a previsão. Cada valor previsto, é incorporado no *input* da previsão da hora seguinte, apresentando assim um comportamento de janela deslizante.

De modo a realizar a previsão para as próximas 24 horas de cada uma das *features* apresentadas como essenciais, foram modeladas 3 redes com a seguinte constituição.

| Model: "sequential" | | |
|---------------------|------------------|---------|
| Layer (type) | Output Shape | Param # |
| lstm (LSTM) | (None, 168, 64) | 16896 |
| lstm_1 (LSTM) | (None, 168, 128) | 98816 |
| lstm_2 (LSTM) | (None, 168, 128) | 131584 |
| lstm_3 (LSTM) | (None, 128) | 131584 |
| dense (Dense) | (None, 64) | 8256 |
| dropout (Dropout) | (None, 64) | 0 |
| dense_1 (Dense) | (None, 1) | 65 |

Figura 4: Arquitetura das redes utilizadas para previsão

6.1 Avaliação da eficiência do ajuste

Como forma de verificar a eficiência do ajuste realizado pelo modelo aos dados reais, são apresentados os seguintes gráficos com ambos os dados. Juntamente, são apresentadas as métricas RMSE (Root Mean Square Error) e MAE (Mean Absolute Error) de modo a avaliar a eficiência do ajuste.

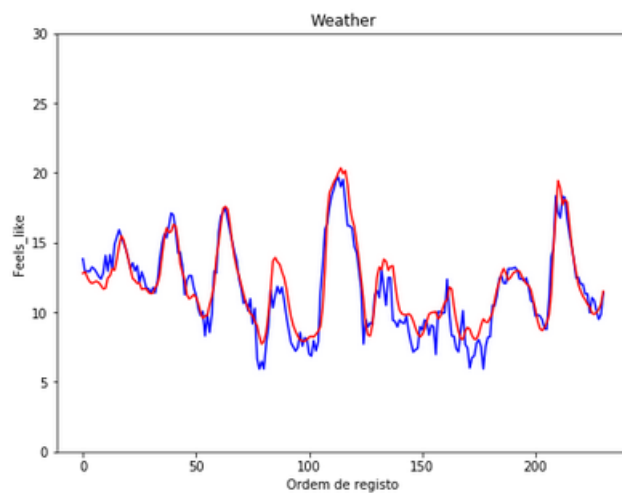


Figura 5: Ajuste realizado pelo modelo aos dados reais da temperatura

- RMSE: 1.26
- MAE: 0.97

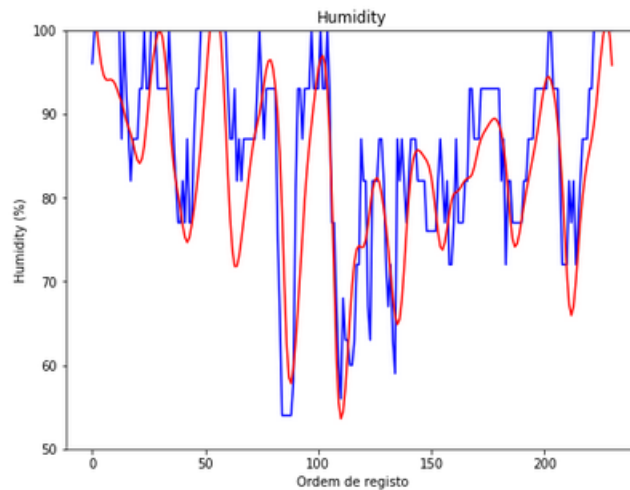


Figura 6: Ajuste realizado pelo modelo aos dados reais da humidade

- RMSE: 7.52
- MAE: 5.78

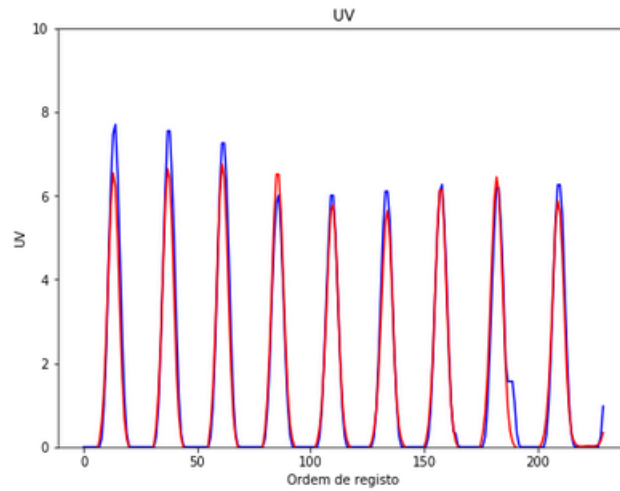


Figura 7: Ajuste realizado pelo modelo aos dados reais do índice ultra-violeta

- RMSE: 0.42
- MAE: 0.27

Nas figuras apresentadas, a eficiência da rede é testado contra os dados com que esta foi treinada. Desta forma, conseguimos avaliar até que ponto a rede aprendeu com os dados que foram utilizados no seu treino. No entanto, esta avaliação é ingênua, não nos garantido certezas sobre a qualidade da previsão de dados futuros.

6.2 Previsão vs Realidade

Antes da previsão final para as próximas 24 horas, foi realizada uma previsão para um dia com os dados já obtidos pelos sensores, de modo a comparar a eficácia das previsões realizadas para esse dia.

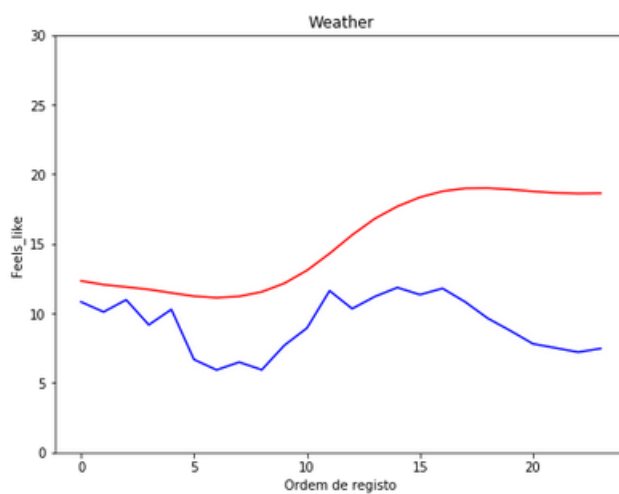


Figura 8: Previsão realizada da temperatura pelo modelo para dados históricos

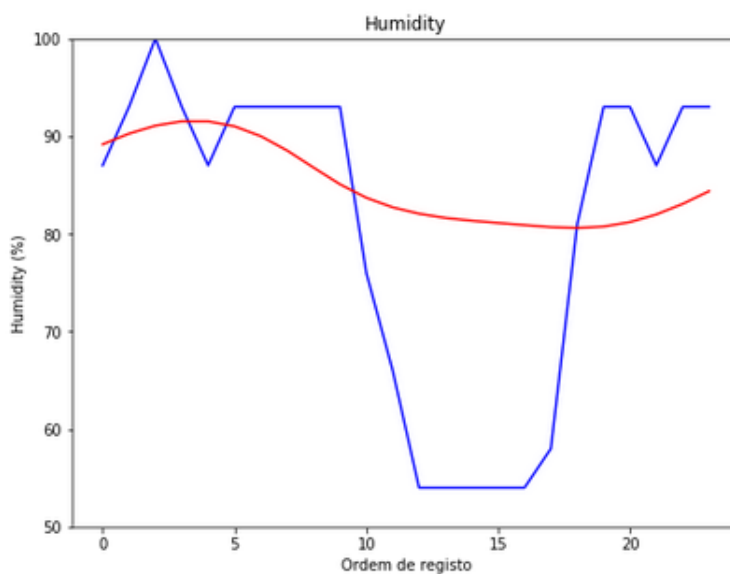


Figura 9: Previsão realizada da humidade pelo modelo para dados históricos

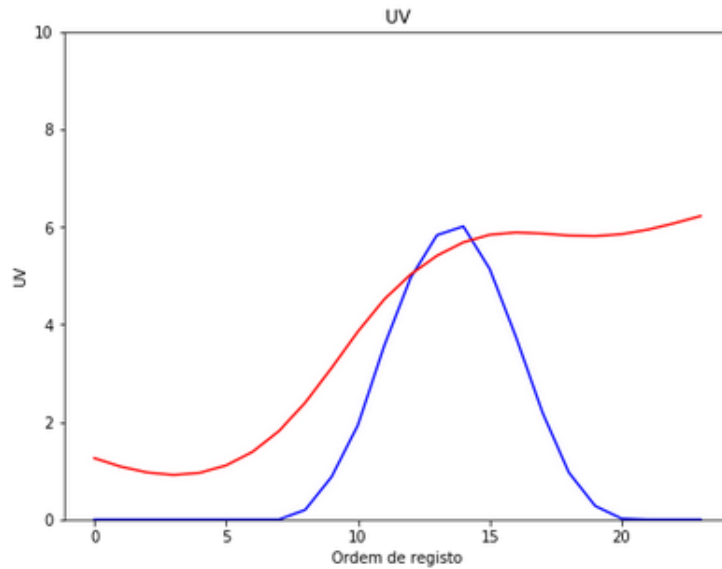


Figura 10: Previsão realizada do índice ultra-violeta pelo modelo para dados históricos

Analisando os resultados obtidos, podemos verificar que os valores previstos pelas redes conseguiram aproximar-se dos valores que foram efetivamente registados. No entanto, no caso da *feature humidity*, a rede não conseguiu prever a queda acentuada da percentagem de humidade entre as 12h e as 16h. No caso apresentado, a rede que aparenta estar mais longe da realidade é a relativa ao índice ultra-violeta, que prevê um valor ultra-violeta elevado para o período noturno (20h às 24h). Isto pode ser devido ao facto de as redes terem sido treinadas apenas os valores registados, não sendo disponibilizado nenhum contexto sobre o momento em que se registou um determinado valor. Se por exemplo, cada valor fosse acompanhado da hora a que foi registado, a rede poderia identificar uma correlação forte entre este par, pelo que quanto mais elevada fosse a hora menor seria o valor do índice ultra-violeta.

6.3 Previsão - 24 horas

Após efetuado o treino da rede, foram realizadas as previsões para as próximas 24 horas para cada uma das *features*.

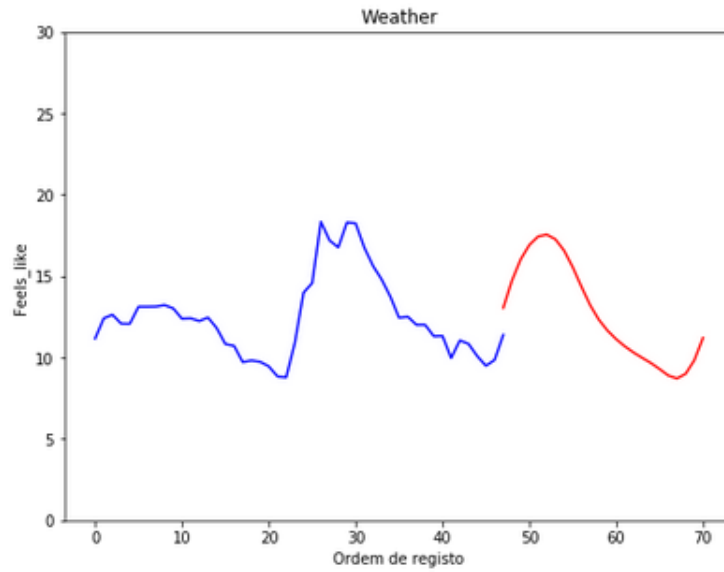


Figura 11: Previsão da temperatura para as próximas 24 horas

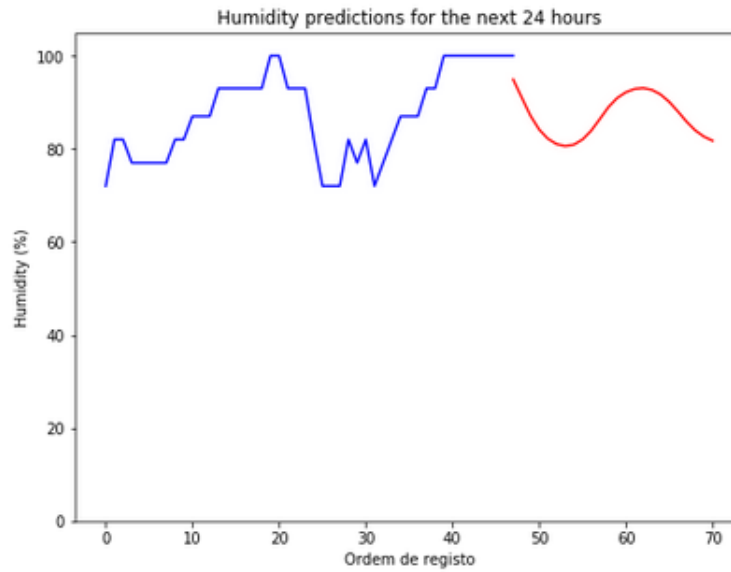


Figura 12: Previsão da humidade para as próximas 24 horas

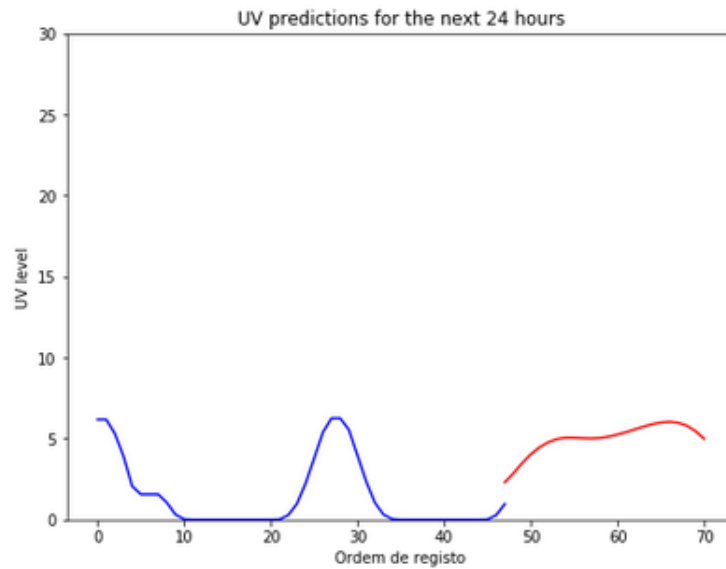


Figura 13: Previsão do índice ultra-violeta para as próximas 24 horas

6.4 Classificação e armazenamento

Após realização das previsões, é executada uma função que atribui uma classificação a cada previsão tendo em conta a temperatura, a humidade e o índice ultravioleta. Isto permitirá ter uma visão mais clara sobre as horas que serão mais apropriadas para a realização de atividades físicas.

Finalmente, toda esta informação, à semelhança do processo de consulta às APIs de dados, é enviada para o backend para posteriormente ser consultada pelo frontend.

7 Interface do sistema

De modo a disponibilizar a informação recolhida pelos sensores e as previsões realizadas pelos modelos, foi criada uma interface em *React*. Esta apresenta-se como interativa e objetiva, facilitando por isso a análise dos dados do sistema. Esta interface, constituinte do front-end, possui conexão direta à base de dados do *Cloud Firestore*. Através destes dois módulos, o sistema apresenta, de uma forma processada, os dados inicialmente brutos recolhidos pelo sensor.

A interface foi dividida em três páginas principais, sendo elas:

- Dados sensorizados do clima;
- Dados sensorizados da radiação ultravioleta;
- Previsão das melhores horas para prática de desporto.

Nesta secção vão ser apresentadas as principais funcionalidades de cada página.

7.1 Dados sensorizados do clima

Esta primeira página encontra-se dividida em dois blocos. O bloco de cima apresenta a temperatura efetiva e a temperatura sentida de hora em hora, em relação às últimas 24 horas.

No bloco de baixo são apresentados as condições das mesmas 24 horas, descrevendo assim mais detalhadamente informações como a condição geral do tempo, a percentagem da humidade, a velocidade do vento sentido e a direção desse mesmo vento.

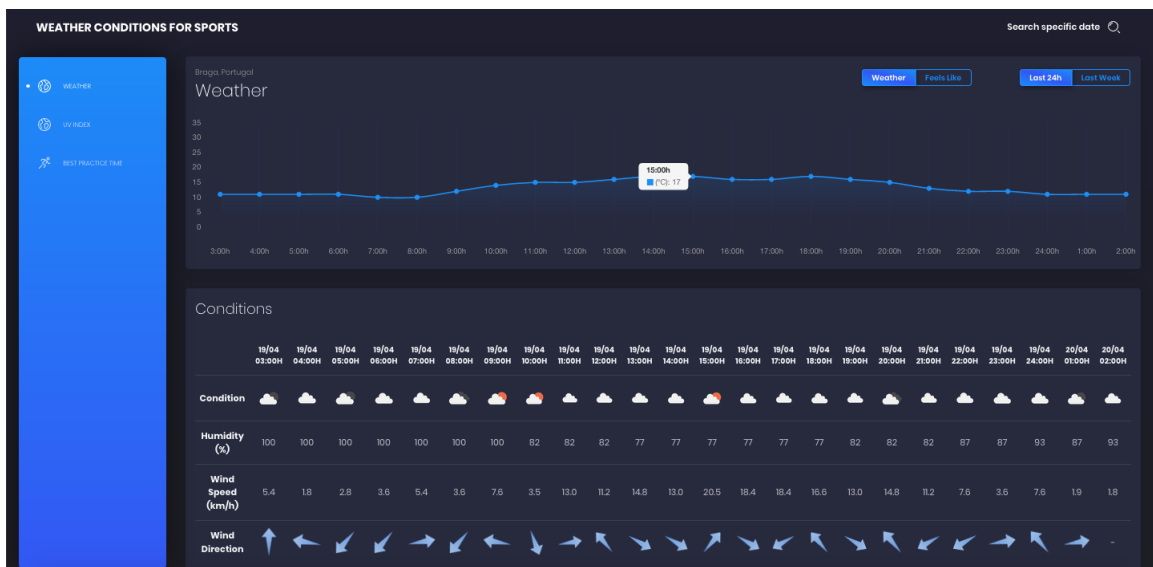


Figura 14: Visão geral da primeira página

7.1 Dados sensorizados do clima

Além destas funcionalidades, é também possível procurar os dados sensorizados em datas anteriores, recorrendo ao calendário associado ao símbolo de procura disponibilizado no canto superior direito.

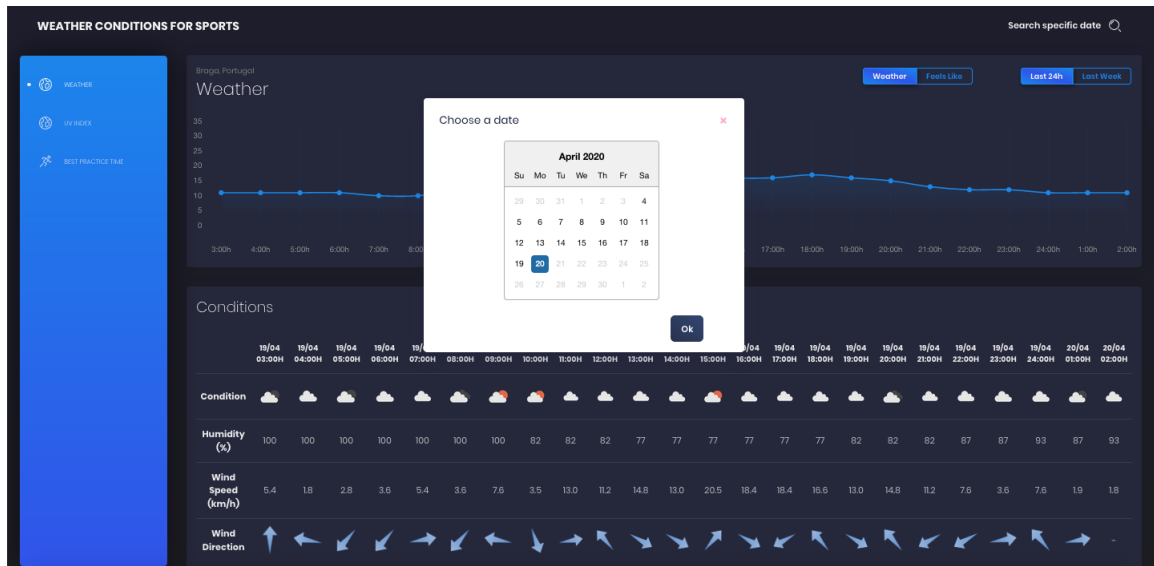


Figura 15: Procura por dados sensorizados noutras datas

7.2 Dados sensorizados da radiação ultravioleta

A segunda página encontra-se dividida em três blocos. O bloco superior apresenta o índice de radiação ultravioleta sentido de hora em hora, em relação às últimas 24 horas, e apresenta também o máximo de índice atingido nos últimos 7 dias, de forma a ser possível ter uma noção dessa variação durante a semana.

No bloco de baixo do lado esquerdo, são apresentados os tempos de exposição solar segura em minutos por cada tipo de pelo, de acordo com a escala de *Fitzpatrick*. No bloco do lado direito é apresentada essa mesma escala, assim como os seus seis tipos e as principais características de cada um para que o utilizador possa identificar o seu tipo de pele, e desta forma realizar uma análise mais pessoal dos dados inicialmente sensorizados.



Figura 16: Visão geral da segunda página

7.3 Previsão das melhores horas para prática de desporto

A terceira página encontra-se também dividida em três blocos. No bloco superior são apresentados os resultados da previsão realizada pelo modelo da secção anterior. Nesta é possível visualizar os dados da temperatura, ou da humidade, ou do índice UV, qualquer um dos três relativos às 24 horas seguintes.

Nos blocos de baixo são apresentados os momentos recomendados e aceitáveis para a prática de desporto das 24 horas seguintes. Estes dados são uma reflexão originada da previsão realizada com o modelo em conjunto com as condições ideais conhecidas para prática de desporto, apresentadas na fundamentação teórica.

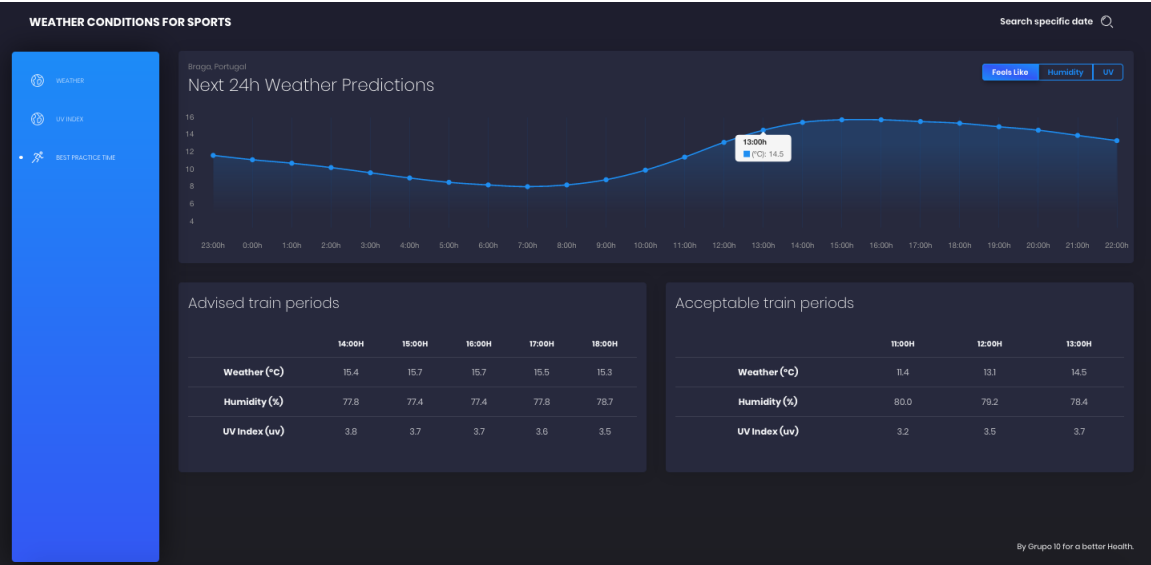


Figura 17: Visão geral da terceira página

8 Conclusão

Dado o trabalho como concluído, o grupo considera que fez, no geral, um bom trabalho. Consideramos que os pontos mais importantes foram alcançados com distinção, especialmente a fase de sensorização. Mesmo naqueles que se calhar não seriam tão relevantes, o grupo esforçou-se no sentido de utilizar ferramentas novas e, conseqüentemente, aprender novos conceitos. A utilização do *Firebase* para armazenar dados e a utilização do *Compute Engine* da *Google Cloud Platform* para correr o *script* de recolha de dados num intervalo regular de tempo, são exemplos disso.

Tivemos alguns problemas na fundamentação do problema porque não existe uma grande amostra de estudos certificados, nem de investigação no que diz respeito às condições ideais para praticar exercício físico. Para conseguirmos tomar decisões relativamente a isto, tivemos que optar por utilizar o conhecimento do grupo nestas matérias e utilizar o conhecimento extraído dos escassos artigos nesta área. Felizmente, não nos causou grande embaraço já que o foco principal não era propriamente na classificação mas sim mais na previsão.

Relativamente à utilização dos dados sensorizados no sistema, importante referir que, apesar da recolha de dados do *Open AQ* ser na mesma efetuada, estes não sofreram quase nenhuma variação durante o período de recolha. Devido a tal, o grupo não achou muito interessante realizar a apresentação destes na interface gráfica.

A previsão de valores foi uma novidade e, naturalmente, trouxe alguns problemas especialmente na fase de compreensão e implementação. Realmente, as **LSTMs** foram um desafio, mas acabaram por nos acrescentar imenso em termos de conhecimento para desafios futuros.

Concluindo, infelizmente não foi possível utilizar sensores reais devido à situação atual do país e do mundo. De qualquer forma, sentimos que estaríamos prontos a aceitar o desafio caso ele surgisse, já que este projeto aumentou exponencialmente a nossa capacidade teórica e prática em assuntos de sensorização.

Referências

- [1] El Helou N, Tafflet M, Berthelot G, et al. Impact of environmental parameters on marathon running performance. PLoS One. 2012;7(5):e37407. doi:10.1371/journal.pone.0037407
- [2] Intermountain Healthcare. “What Is the Effect of Heat and Humidity on Athletic Performance.” Intermountainhealthcare.Org, 21 June 2014, intermountainhealthcare.org/blogs/topics/sports-medicine/2014/06/what-is-the-effect-of-heat-and-humidity-on-athletic-performance/.
- [3] Zhao, Jiexiu, et al. “Effects of Heat and Different Humidity Levels on Aerobic and Anaerobic Exercise Performance in Athletes.” Journal of Exercise Science Fitness, vol. 11, no. 1, June 2013, pp. 35–41
- [4] “How to Calculate the Effect of Humidity on Running Performance - Runners Connect.” Runners Connect, 15 Aug. 2016, runnersconnect.net/dew-point-effect-running
- [5] Casanova, Amanda. “Running in the Humidity.” ACTIVE.Com, 7 May 2015, www.active.com/running/articles/running-in-the-humidity.
- [6] Índice ultravioleta, https://pt.wikipedia.org/wiki/Índice_ultravioleta