



**Politécnico
de Viseu**

Escola Superior
de Tecnologia
e Gestão de Viseu

Modelo Preditivo de Estimativa do Peso Corporal Baseado em Inteligência Artificial: Uma Abordagem Integrada de Pré-processamento e Avaliação

Diana Margarida Lopes Figueiredo

Dissertação

Mestrado em Engenharia Informática - Sistemas de Informação

Trabalho efetuado sob a orientação de
Professor Doutor Carlos Augusto da Silva Cunha
Professor Doutor Rui Pedro Monteiro Amaro Duarte

Março de 2024



**Politécnico
de Viseu**

Escola Superior
de Tecnologia
e Gestão de Viseu

Modelo Preditivo de Estimativa do Peso Corporal Baseado em Inteligência Artificial: Uma Abordagem Integrada de Pré-processamento e Avaliação

Diana Margarida Lopes Figueiredo

Dissertação

Engenharia Informática -Sistemas de Informação

Trabalho efetuado sob a orientação de

Professor Doutor Carlos Augusto da Silva Cunha
Professor Doutor Rui Pedro Monteiro Amaro Duarte

Março de 2024

Agradecimentos

Um enorme agradecimento para o professor Doutor Carlos Augusto da Silva Cunha e para o professor Doutor Rui Pedro Monteiro Amaro Duarte, por toda a ajuda, aconselhamento e tempo dispensado.

Um muito obrigado aos meus pais, por todo o apoio que sempre me deram e por me fornecerem a possibilidade de atingir os meus objetivos. Por fim ao meu colega, Rui Jorge Marques Oliveira, por toda a amizade e apoio.

Muito obrigado a todos.

Resumo

O peso corporal é muito mais que um número numa balança. Este valor pode ser indicador de várias doenças, pois tanto o excesso como a falta de peso têm implicações na saúde dos indivíduos. O excesso de peso está associado a doenças cardíacas, obesidade, diabetes, hipertensão arterial, distúrbios respiratórios, entre outras. Enquanto a falta de peso num nível extremo está associado a problemas de deficiência nutricional, enfraquecimento do sistema imunológico, osteoporose e desequilíbrios hormonais. Devido a estes problemas, surge a necessidade de acompanhar e analisar as alterações corporais, para a adoção de uma dieta e um estilo de vida equilibrado com as necessidades do indivíduo.

O processo de controlo do peso é um processo complicado e está dependente de vários fatores. Assim sendo, e considerando que a versatilidade da área de *Machine Learning* (ML) permite desenvolver projetos que melhorem a qualidade de vida do ser humano, neste trabalho pretende-se desenvolver um modelo de ML para prever o peso futuro tendo em conta o registo alimentar, exercício físico e *Taxa Metabólica Basal* (TMB) de indivíduo, com o objetivo de mostrar o impacto que três dias podem ter no peso futuro.

Os resultados da performance do modelo obtidos através do cálculo das métricas de desempenho, foram positivos. Através do cálculo do Coeficiente de Determinação foi obtido o valor 0.75, o que para esta métrica é considerado um valor bom, visto que está mais próximo de 1 do que de 0. Os valores do cálculo do *Mean Squared Error* (MSE) e do *Mean Absolute Error* (MAE) demonstra que o modelo conseguiu aprender os padrões nos dados e que não existiu *overfitting* significativo. Estes resultados demonstram ser viável o desenvolvimento deste tipo de soluções.

Palavras-Chave: Machine Learning, Redes neuronais, Peso Corporal, Taxa metabólica basal, Equação de *Harris e Bennedit*

Abstract

Body weight is much more than just a number on a scale. This value can be indicative of various diseases, as both excess and insufficient weight have implications for individual's health. Excess weight is associated with heart disease, obesity, diabetes, high blood pressure, respiratory disorders, among others. Meanwhile, extreme underweight is associated with problems such as nutritional deficiency, weakened immune system, osteoporosis, and hormonal imbalances. Due to these issues, there is a need to monitor and analyze body changes in order to adopt a diet and lifestyle balanced with individual needs.

The process of weight control is a complicated one and depends on various factors. Therefore, considering the versatility of the field of ML, which allows for the development of projects that improve human quality of life, this study aims to develop a ML model to predict future weight based on dietary records, physical exercise, and TMB, with the goal of demonstrating the impact that three days can have on future weight.

The results of the model's performance obtained through the calculation of performance metrics were positive. The coefficient of determination yielded a value of 0.75, which for this metric is considered good, as it is closer to 1 than to 0. The values of MSE and MAE demonstrate that the model was able to learn patterns in the data and that there was no significant overfitting. These results indicate the viability of developing such solutions.

Keywords: Machine learning, Body Weight, Basal Metabolic Rate, *Harris e Bennedit* Equation, Neural network

Índice

Lista de Tabelas	vii
Lista de Figuras	ix
Lista de Acrónimos	xi
1 Introdução	1
1.1 Motivação	1
1.2 Contextualização	2
1.3 Definição do Problema	4
1.4 Questões de investigação	4
1.5 Objetivos	5
1.6 Resultados Esperados	5
1.7 Plano de Trabalhos	6
1.8 Estrutura da Tese	7
2 Revisão da Literatura	9
2.1 Fundamentos Teóricos	9
2.1.1 Modelos de Machine Learning	9
2.1.2 Avaliação de Modelos	16
2.1.3 Problemas comuns associados aos modelos.	16
2.2 Impacto das alterações de peso nas pessoas	17
2.2.1 Fatores associados	19
2.2.2 Acompanhamento das Alterações	20
2.2.3 Métricas utilizadas na área da nutrição.	21
2.3 Trabalhos Relacionados	22
2.4 Conclusão	25
3 Metodologia	27
3.1 Compreensão do negócio	28
3.2 Compreensão dos dados	29
3.3 Preparação dos dados	29
3.4 Modelação	30
3.5 Avaliação	31

3.6	Considerações Finais	31
4	Implementação	33
4.1	Compreensão dos dados	33
4.2	Preparação dos dados	34
4.3	Modelação	37
4.4	Avaliação	39
4.5	Generalização do modelo	39
4.6	Erro de previsão	40
4.7	Discussão de resultados	42
5	Conclusão	45
5.1	Limitações e desafios	46
5.2	Trabalho futuro	46
	Referências	48

Lista de Tabelas

4.1	Definição dos atributos do <i>dataset</i>	34
4.2	Detalhes dos Dados escolhidos	35
4.3	Valores dos Hiperparâmetros.	38
4.4	Valores das variáveis Y_{pred} e Y_{test}	42

Lista de Figuras

1.1	Plano de Trabalhos proposto.	7
2.1	Arquitetura da MLP	12
2.2	Funções linear e não linear	13
2.3	Função <i>Rectified Linear Units</i> (ReLU) [Zhang et al., 2023]	14
2.4	Cenário de <i>underfitting</i> e <i>overfitting</i> [geeksforgeeks, 2023]	17
3.1	Diagrama CRISP-DM Fonte:[Chapman et al., 2000]	28
3.2	Diagrama do processo de preparação dos dados.	30
4.1	Tabela com os dados após o pré-processamento	35
4.2	Tabela com o cálculo da taxa metabólica basal (TMB)	36
4.3	Tabela com o cálculo do peso futuro tendo em conta os últimos três dias	37
4.4	Diagrama representativo do modelo de rede neuronal implementado	38
4.5	Diagrama do processo de treino e de teste do modelo	39
4.6	Gráfico do cálculo do MSE e do MAE por cada época	40
4.7	Gráfico da comparação entre o peso atual e o peso previsto	41

Lista de Acrónimos

CNN	<i>Convolutional Neural Network</i>
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
CSV	<i>Common Separated Values</i>
DL	<i>Deep Learning</i>
IMC	<i>Índice de Massa Corporal</i>
LR	<i>Logistic Regression</i>
LSTM	<i>Long Short-Term Memory</i>
MAE	<i>Mean Absolute Error</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Squared Error</i>
ReLU	<i>Rectified Linear Units</i>
RF	<i>Random Forest</i>
RMSprop	<i>Root Mean Square Propagation</i>
RNA	<i>Redes neuronais artificiais</i>
RNN	<i>Recurrent Neural Networks</i>
TMB	<i>Taxa Metabólica Basal</i>

Capítulo 1

Introdução

O peso corporal humano é utilizado na área da saúde e da nutrição como um indicador da saúde e bem-estar [Svendsen et al., 2020]. O acompanhamento da alteração do peso corporal faz com que seja possível diminuir o risco de doenças relacionadas com uma alimentação não saudável [Blüher, 2020]. Desta forma existe um entendimento das diferenças nas trajetórias de ganho de peso e a capacidade de prever o peso futuro têm implicações significativas para a promoção de uma vida saudável e para o desenvolvimento de estratégias de intervenção personalizadas, na medida em que previnem doenças relacionadas com um estilo de vida menos saudável. Este estudo propõe a pesquisa e desenvolvimento de um modelo de *Machine Learning* (ML) para analisar as diferenças de peso corporal de um humano e prever o seu peso futuro, visando facilitar a compreensão do impacto que a alimentação e o exercício físico têm na alteração da massa corporal.

1.1 Motivação

O peso corporal, normalmente interpretado pelas pessoas como um número na balança, mas é muito mais do que isso. Este indicador tem uma grande importância na vida do ser humano, que afeta não só a aparência estética, mas também impacta na sua saúde física [Abhari et al., 2019].

O excesso e a falta de peso têm implicações na saúde dos indivíduos. O excesso de peso aumenta o risco de aparecimento de doenças cardíacas, obesidade, diabetes, hipertensão arterial, distúrbios respiratórios, entre outras [Kolotkin and Andersen,

2017]. Contrariamente, os indivíduos que têm falta de peso sofrem de problemas associados a deficiência nutricionais, enfraquecimento do sistema imunológico, osteoporose e desequilíbrios hormonais [Le et al., 2017]. Para além disto, não é apenas a saúde física que sofre consequências das alterações de peso, mas também a saúde psicológica é afetada [Moradi et al., 2021]. Muitas vezes, a pressão social que é imposta para cumprir padrões estéticos difíceis de manter e de alcançar acabam por impactar na autoestima de quem sofre de excesso ou diminuição excessiva de peso. O aparecimento destas doenças impacta negativamente a saúde das pessoas, limitando a mobilidade, causando fadiga e impactando a qualidade de vida em geral [Marengo et al., 2018]. Para as combater, é necessário elaborar uma dieta adequada as necessidades nutricionais de cada pessoa. A introdução de uma vida ativa, acompanhamento profissional e por fim acompanhar as alterações de peso para ver se estas medidas surgem efeito são algumas das soluções para prevenir estes problemas.

O controlo de peso é uma jornada pessoal e desafiadora para muitos indivíduos, uma vez que, por vezes, obriga a introduzir mudanças importantes no seu estilo de vida como, por exemplo, adaptar os planos alimentares a diferentes pesos corporais, controlar as alterações de peso e adaptar o estilo de vida as diferentes faixas de peso [Cheong and Freene, 2018]. Assim, a grande motivação deste trabalho é compreender o impacto que as alterações de peso representam na vida do ser humano, o que possibilita projetar intervenções que incentivem melhorias no estilo de vida e reduzam o risco de serem afetados por doenças derivadas de uma alimentação pobre em nutrientes e de uma vida sedentária.

1.2 Contextualização

Um dos pilares fundamentais da ciência da saúde é a investigação dos determinantes da saúde, ou seja, os fatores que contribuem positivamente ou negativamente para o estado de saúde de um indivíduo ou da população [Cockerham et al., 2017]. Isso inclui não apenas os aspetos biológicos, como a genética e predisposição a doenças, mas também fatores comportamentais, sociais e económicos [Albuquerque et al., 2017]. A prevenção de doenças é abordada de maneira pro-ativa, visando a identificação e mitigação dos fatores de risco antes que eles se manifestem como problemas de saúde [Garrard, 2020]. Além de focar na prevenção, a ciência da saúde dedica-se à promoção de hábitos saudáveis. Isso vai além do simples tratamento de doenças e envolve a criação de ambientes propícios à adoção de comportamentos saudáveis [Xuan et al., 2020].

A influência direta da nutrição no peso corporal e, conseqüentemente, na saúde geral é destacada na área da ciência da saúde. Os hábitos alimentares, escolhas nutricionais e a qualidade da dieta são elementos cruciais na equação que molda o peso de um indivíduo [Harrop et al., 2021].

A nutrição, como componente essencial da saúde, é responsável pela manutenção das funções fisiológicas, ajuda na prevenção de doenças e melhora a qualidade de vida e o bem-estar [Brannon et al., 2018]. O consumo equilibrado de nutrientes é crucial para o desenvolvimento e funcionamento adequado do organismo, afetando desde os níveis celulares até a saúde global do indivíduo [Carnauba et al., 2017]. O impacto da nutrição transcende as fronteiras da biologia, influenciando aspectos psicológicos e sociais da vida humana.

A obesidade é uma dessas doenças e tornou-se uma preocupação de saúde pública global. Segundo dados da Organização Mundial da Saúde, estima-se que em 2022 mais de 1,9 bilhão de adultos em todo o mundo estivessem com excesso de peso, e mais de 650 milhões eram classificados como obesos [Horta et al., 2023]. Esses números alarmantes refletem não apenas uma mudança nos padrões alimentares, mas também um estilo de vida sedentário, fatores socioeconômicos que contribuem para uma alimentação pobre em nutrientes e rica em alimentos processados que originam um acúmulo de excesso de peso [Buoncristiano et al., 2021].

A relação entre a obesidade e as doenças cardiovasculares é inegável. Estudos demonstram que a obesidade está associada a um aumento significativo no risco de desenvolver hipertensão, diabetes, doenças respiratórias, doenças cardíacas, entre outras [Powell-Wiley et al., 2021]. Conforme a *World Heart Federation*, aproximadamente 2,8 milhões de mortes por ano são atribuídas à obesidade, mostrando assim a urgência de abordagens preventivas [Lopez-Jimenez et al., 2022].

Acredita-se que os sistemas de previsão sejam uma solução para a prevenção de doenças associadas ao aumento do peso e assim controlar este fator que tanto impacta a saúde humana [Safaei et al., 2021]. Nos últimos anos foram desenvolvidos vários estudos em relação a sistemas de previsão do impacto que o estilo de vida tem na saúde, no peso corporal ou previsão do aumento de risco de contrair doenças como a obesidade, hipertensão ou outro tipo de problemas [Zheng and Ruggiero, 2017]. Este tema é bastante comum na comunidade científica devido à importância que a saúde tem na vida do ser humano e a necessidade de melhorar a vida das pessoas. Assim, um sistema de previsão de peso futuro é um sistema que tem em conta fatores que influenciam o peso como, por exemplo, as calorias ingeridas, a prática diária de exercício físico diariamente e o tipo de alimentação que uma pessoa tem.

Para a previsão do peso futuro, pode-se encontrar vantagens na utilização de dados históricos de forma mais eficaz. Os modelos de ML e nos seus sub-campos têm-se mostrado proficientes em analisar grandes conjuntos de dados longitudinais, identificando padrões e correlações que escapariam às análises convencionais [Alkhalaf et al., 2022]. Essa análise retrospectiva permite prever mudanças no peso corporal com uma precisão cada vez maior. Outro avanço crucial é a incorporação de variáveis relevantes na construção de modelos preditivos. Em contraste com abordagens

tradicionais que se baseiam em fatores isolados, os modelos de ML conseguem integrar uma variedade de variáveis, desde dados históricos de atividade física até a hábitos alimentares [Zhao et al., 2021]. Essa abordagem mais abrangente permite uma análise mais precisa e personalizada, considerando a singularidade de cada indivíduo.

Apesar dos progressos notáveis, a utilização de técnicas de ML na análise do peso corporal ainda enfrentam desafios, como a necessidade de dados de alta qualidade, questões de privacidade dos dados e a interpretabilidade dos modelos. No entanto, esses desafios abrem oportunidades para futuras pesquisas, destacando a importância de abordagens mais éticas, mais detalhadas e individualizadas.

1.3 Definição do Problema

Tal como referido nas secções anteriores, o peso corporal humano constitui um parâmetro muito importante no contexto da saúde e nutrição, sendo amplamente empregado como indicador nesses domínios. A monitorização das oscilações no peso corporal desempenha um papel crucial na mitigação do risco de patologias associadas a uma alimentação desequilibrada. A compreensão das disparidades nas trajetórias de aumento de peso e a capacidade de antecipar o peso futuro revelam-se elementos fundamentais na promoção de estilos de vida saudáveis e na definição de estratégias de intervenção personalizadas.

Os hábitos alimentares saudáveis e a nutrição adequada desempenham um papel de relevo no processo de desenvolvimento humano. O controlo do peso constitui uma jornada desafiadora para muitos, exigindo alterações substanciais no estilo de vida, tais como a adoção de uma alimentação mais equilibrada, a redução da ingestão calórica e o aumento da atividade física. A compreensão do impacto destes fatores na existência humana reveste-se de importância crucial para o desenho de intervenções que promovam melhorias no estilo de vida

Contudo, a adesão a planos alimentares saudáveis pode revelar-se desafiadora em virtude de diversos fatores, tais como a complexidade do acompanhamento nutricional e influências externas, como a saúde psicológica, fatores económicos e sociais. Neste contexto, este projeto visa compreender as razões para a baixa adesão a planos alimentares e desenvolver estratégias que incentivem os indivíduos a manterem os seus planos alimentares, promovendo assim um estilo de vida saudável e ativo.

1.4 Questões de investigação

Considerando o referido nas secções anteriores, as perguntas a que se pretende responder com a realização deste trabalho são:

- *(Q1)*: É viável desenvolver modelos preditivos capazes de estimar de forma precisa e imediata as consequências da ingestão de calorias e da prática de exercício físico no peso corporal no curto prazo?
- *(Q2)*: Que técnicas de pré-processamento de dados devem ser consideradas para melhorar a qualidade e eficácia dos modelos preditivos que procuram estimar as consequências imediatas da ingestão de calorias e da prática de exercício físico no peso corporal, incluindo fatores como normalização, tratamento de dados em falta, seleção de características relevantes?
- *(Q3)*: Qual é o desempenho do modelo de previsão do peso corporal tendo em conta a ingestão de calorias e a prática de exercício físico?

1.5 Objetivos

Este projeto contempla o desenvolvimento de modelos de ML para previsão do peso futuro tendo em conta as calorias ingeridas e gastas de um determinado indivíduo durante um determinado espaço temporal. Para tal é necessário:

- *O1*: desenvolver modelo preditivo de estimativa precisa e imediata das consequências da ingestão de calorias e da prática de exercício físico no peso corporal, utilizando técnicas ML. Este objetivo visa fornecer uma ferramenta eficaz para prever mudanças no peso corporal com base em dados históricos e variáveis relevantes.
- *O2*: identificar e implementar técnicas de pré-processamento de dados para melhorar a qualidade e eficácia dos modelos preditivos. Isso inclui a normalização, tratamento de dados em falta, seleção de características relevantes, visando otimizar a precisão e a capacidade de generalização dos modelos.
- *O3*: avaliar a precisão da previsão do peso corporal com base no modelo desenvolvido. Isso envolverá a utilização de métodos como métricas de avaliação e seleção de características, visando compreender melhor a eficácia do modelo em prever as mudanças no peso corporal e identificar possíveis áreas de melhoria.

1.6 Resultados Esperados

Com base nos objetivos identificados na secção anterior, espera-se atingir os seguintes resultados:

- *(R1)*: Implementação de uma solução de previsão baseada em modelos ML: o peso é um fator importante na vida das pessoas, não só a nível estético

mas também a nível de saúde. Assim, considerando os fatores que podem influenciar o peso corporal, será implementado um modelo de ML que irá prever o impacto que o consumo e o gasto de caloria tem no peso futuro.

- (*R2*): Identificação e implementação de técnicas de pré-processamento de dados que melhorem a qualidade e eficácia dos modelos de ML. Espera-se que essas técnicas, como normalização, tratamento de dados, seleção de características relevantes, contribuam para aumentar a precisão e a capacidade de previsão do modelo;
- (*R3*): Obter resultados o mais próximos da realidade possíveis. Esse modelo será capaz de estimar com acuidade as consequências imediatas da ingestão de calorias e da prática de exercício físico no peso corporal, proporcionando *insights* valiosos para a promoção de estilos de vida saudáveis e intervenções personalizadas na área da saúde e nutrição.

De uma forma geral, os resultados esperados devem incluir um modelo preditivo preciso e eficaz, técnicas de pré-processamento de dados bem-sucedidas e uma compreensão aprofundada da precisão e eficácia do modelo em prever as mudanças no peso corporal. Esses resultados têm o potencial de contribuir significativamente para a promoção de estilos de vida saudáveis e para o desenvolvimento de intervenções personalizadas na área da saúde e nutrição.

1.7 Plano de Trabalhos

Para atingir os objetivos, com foco nos resultados esperados, foram definidas as seguintes macro-etapas para o projeto. Estas macro-etapas estão representadas na Figura 1.1.

Numa primeira fase, foi analisada a influência do peso na saúde através da revisão da literatura com ênfase na forma como o fator peso é utilizado na medicina e na nutrição, na perspectiva da forma que este fator impacta a vida das pessoas e a importância de um acompanhamento regular das alterações do peso.

A seguir, foi realizada uma análise de métodos de ML que poderão ser aplicados no contexto da previsão do peso futuro. De forma a analisar o que já existe dentro da área do desenvolvimento de modelos de ML para a área da nutrição, foi necessário realizar uma pesquisa de trabalho relacionado, onde se irá proceder à pesquisa de trabalhos e aplicações que poderão servir de apoio ao desenvolvimento/estudo deste projeto.

Na implementação do projeto, foi necessário proceder à recolha de dados e preparação dos mesmos. Nesta fase do projeto, foram identificados conjunto de dados para usar no desenvolvimento do modelo de ML. Neste contexto, a qualidade e a quantidade das informações dos dados são importantes para um melhor desempenho

do modelo de ML. Depois da recolha de dados foi possível iniciar desenvolvimento do modelo de ML. Na bibliografia analisada foi possível identificar o melhor método de ML a aplicar à modelação dos dados.

Terminada a fase de modelação procedeu-se à fase de treino e de teste do modelo. Treinar e testar o modelo envolvendo a seleção de atributos e experimentar com diversos hiperparâmetros é fundamental para obter o melhor resultado possível. A ideia subjacente é fornecer dados ao modelo, permitindo-lhe aprender e aprimorar o seu desempenho. Após este processo de treino, o modelo adquire um conhecimento mais abrangente, aumentando assim a sua capacidade de realizar previsões sobre o peso futuro tendo em conta as variáveis de entrada.

Finalmente, seguiu-se a fase de avaliação dos resultados que permitiu avaliar o desempenho do modelo. Essa avaliação não apenas fornece *insights* sobre o quão eficaz o modelo é inicialmente, mas também oferece orientação para possíveis ajustes que podem ser feitos para aprimorar o seu desempenho.

Por fim, termina-se com a escrita deste documento que regista todas as aprendizagens e os resultados alcançados com a realização deste projeto.

Nome da tarefa	2022				2023												2024			
	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dec	Jan	Fev	Mar	
Estudo da influencia do peso na saude																				
Estudo de métodos de Machine Learning																				
Pesquisa de trabalho relacionado																				
Recolha e preparação de dados																				
Desenvolvimento do modelo de Machine Learning																				
Treinar e testar o modelo																				
Avaliação dos resultados																				
Escrita da dissertação																				

Figura 1.1: Plano de Trabalhos proposto.

1.8 Estrutura da Tese

Esta secção apresenta a estrutura da tese, incluído a organização de cada capítulo e a relação entre eles.

No presente capítulo, é feita uma introdução ao tema abordado de forma a ter um conhecimento geral sobre o tema proposto. É abordado a motivação, contextualização, definição do problema, os objetivos assim como os resultados esperados e o plano de trabalhos. Através da contextualização é possível ter acesso a pequenos excertos da literatura realizada.

Através do Capítulo 2 irá ser realizada a revisão da literatura, sendo abordados temas como a importância do peso na vida do ser humano, fatores que podem influenciar a alteração de peso, a importância de acompanhar as alterações de peso, métodos de cálculo do peso ideal. Foi criada também uma secção com conceitos fundamentais para contextualização de métodos. Nesta secção é possível encontrar informação sobre ML e outros temas relacionados. Foi necessário realizar também pesquisa bibliográfica com ênfase em trabalhos relacionados.

No Capítulo 3 é possível perceber a metodologia utilizada para o desenvolvimento do modelo. Nesta secção é descrito o processo da investigação, incluindo a amostra, métodos de recolha de dados, as técnicas utilizadas no estudo e a justificação da fundamentação dos métodos escolhidos e as suas potenciais limitações.

O Capítulo 4 refere-se ao processo de implementação do modelo com uma descrição dos vários passos realizados e as decisões tomadas. Este capítulo termina com a análise dos resultados da solução implementada.

Por fim, no Capítulo 5 são apresentadas as principais conclusões com ênfase nas limitações e desafios subjacentes ao trabalho, bem como perspectivas sobre trabalho futuro.

Capítulo 2

Revisão da Literatura

Neste capítulo são apresentados os fundamentos teóricos relacionados com impacto que o peso tem na vida das pessoas, acompanhamento das alterações e métodos de cálculo do peso ideais. Será também abordado os trabalhos relacionados com este projeto, e as técnicas de ML que possam ser soluções ao problema.

2.1 Fundamentos Teóricos

Nesta secção vão ser abordados conceitos fundamentais associados ao trabalho proposto nomeadamente no que diz respeito aos modelos de ML. Para estes, são também identificadas as formas de avaliação de modelos e os problemas de *overfitting* e o *underfitting*.

2.1.1 Modelos de Machine Learning

O ML, é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem aos computadores aprender padrões a partir de dados [Badillo et al., 2020a]. Em vez de programar explicitamente instruções para executar uma tarefa específica, os sistemas de ML são projetados para aprender e melhorar com a experiência. O processo de ML envolve a exposição do sistema a um conjunto de dados, no qual padrões ou relações são identificados. Com base nesses padrões aprendidos, o modelo tem a capacidade de fazer previsões ou tomar decisões quando confrontado com novos dados [Molnar, 2020].

As aplicações dos modelos de ML são vastas e abrangem diversas áreas, desde diagnósticos médicos e previsão de mercado até reconhecimento de voz, recomendação de produtos e carros autônomos [Zhou, 2021]. A capacidade de aprender com dados e melhorar o desempenho ao longo do tempo faz do ML uma tecnologia poderosa e inovadora.

Existem diferentes abordagens no ML, sendo as principais a aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço [Morales and Escalante, 2022, Badillo et al., 2020b]. Na aprendizagem supervisionada, o modelo é treinado com um conjunto de dados rotulados, o que significa que cada exemplo de treino é associado a uma saída conhecida. O modelo aprende a relação entre as entradas e as saídas desejadas, permitindo que, posteriormente, faça previsões ou classificações em dados não rotulados [Marsland, 2011]. Exemplos comuns incluem previsão de vendas com base em histórico [Tsoumakas, 2019], diagnóstico médico [Chebli et al., 2018] e reconhecimento de voz [Ravanelli et al., 2020].

No processo de aprendizagem não supervisionado, o modelo é treinado com um conjunto de dados sem rótulos, e o seu objetivo é encontrar padrões e estruturas inerentes nos dados [Sathya et al., 2013]. Isso é frequentemente utilizado nas tarefas de agrupamento, onde o modelo tenta organizar os dados em grupos ou *clusters* com base em características semelhantes. O *k-Means* [Ahmed et al., 2020] é um exemplo clássico desse tipo de algoritmo, sendo usado em diversas áreas tais como segmentação de clientes [Kansal et al., 2018], análise de imagem [Dhanachandra et al., 2015] e reconhecimento de padrões [Peng et al., 2013].

Por fim, a aprendizagem por reforço cujo propósito é aprender consoante o seu meio envolvente e não através de conjuntos de dados inicialmente inseridos. Um exemplo deste método de aprendizagem materializa-se quando um agente interage com um ambiente e é recompensado ou penalizado com base nas suas ações. O agente aprende a realizar ações que maximizem a recompensa ao longo do tempo [Sutton and Barto, 2018]. Esse tipo de aprendizagem é aplicado em jogos [Lample and Chaplot, 2017], controlo de robôs [Johannink et al., 2019] e otimização de processos [Powell et al., 2020].

As *Redes neuronais artificiais* (RNA) são algoritmos computacionais inspirados na estrutura neuronal do cérebro humano e nos últimos anos, as RNA têm emergido como uma ferramenta poderosa e versátil em diversas áreas, desde da visão computacional, previsão do futuro e até ao processamento de linguagens [Dastres and Soori, 2021]. Esta ascensão é impulsionada pela capacidade das redes neuronais de aprenderem padrões complexos e extraírem informações significativas a partir de grandes conjuntos de dados [Bashar, 2019]. Para tal foram desenvolvidos diversos tipos de redes neuronais para atender a diferentes requisitos e desafios como, por exemplo, as *Convolutional Neural Network* (CNN), *Recurrent Neural Networks* (RNN) e as *Multilayer Perceptron* (MLP) [Wu and Feng, 2018].

O primeiro modelo a ser implementado, utilizado esta tecnologia foi chamado de *Perceptron* e foi proposto por Frank Rosenblatt em 1957 [Kanal, 2003]. Este modelo foi concebido com o intuito de simular o funcionamento básico de um neurónio no cérebro humano. O *Perceptron* é uma rede neural de uma única camada, cuja simplicidade proporciona uma compreensão fundamental dos princípios subjacentes às redes neuronais [Rosenblatt, 2021].

As CNN constituem um tipo específico de rede neural desenvolvido para transformar dados provenientes de imagens numa variável de saída. Portanto, no contexto do desenvolvimento de representações internas de imagens 2D, classificação de modelos de previsão e resolução de problemas de previsão de regressão, as CNN demonstram uma grande adequação. De maneira geral, essas redes são especialmente eficazes para lidar com dados que possuem uma relação espacial [Li et al., 2021].

As RNN foram desenvolvidas para lidar com dados sequenciais de maneira eficiente, incorporando uma arquitetura que permite processar e entender padrões temporais nos diferentes conjuntos de dados. A estrutura recursiva da RNN, faz com que os valores da entrada sejam dependentes dos passos que aconteceram nos neurónios anteriores e possibilita que a informação anterior influencie a tomada de decisões nos instantes posteriores, conferindo-lhes uma capacidade única de lidar com a natureza dinâmica e sequencial de muitos tipos de dados [Katte, 2018].

As redes MLP, aplicam-se a classificação de modelos de previsão e a problemas de previsão de regressão, tornaram-se numa das redes mais utilizadas no desenvolvimento de redes neuronais. Normalmente, são utilizadas nos conjuntos de dados em formato tabular e a sua estrutura está dividida numa camada de entrada, uma ou mais camadas ocultas e uma camada de saída, permitindo assim a captura de relações mais complexas nos dados [Almeida, 2020]. A camada de entrada é a primeira camada da MLP sendo composta pelos neurónios que representam as diferentes características dos dados de entrada e os mesmos são denominados unidades de entrada. Cada neurónio nesta camada está associado a uma característica específica e é responsável por distribuir os valores de entrada para as outras camadas sem sofrer alterações. A dimensão dessa camada é determinada pelo número de características nos dados de entrada. As camadas ocultas são responsáveis por processar e transmitir informações por meio das conexões entre as camadas de entrada e saída. Os neurónios pertencentes às camadas intermediárias são denominados unidades ocultas. A profundidade e a quantidade de neurónios nessas camadas impactam diretamente a capacidade da rede de aprender padrões complexos. A camada de saída, contem os neurónios denominados neurónios de saída e responsável por gerar as previsões ou resultados desejados. O número de neurónios nesta camada depende da natureza da tarefa. Por exemplo, em problemas de classificação binária, pode haver um neurónio de saída, enquanto nas tarefas de classificação de multi-classe, haverá um neurónio para cada classe.

Através da Figura 2.1 é possível observar um exemplo da arquitetura de uma MLP.

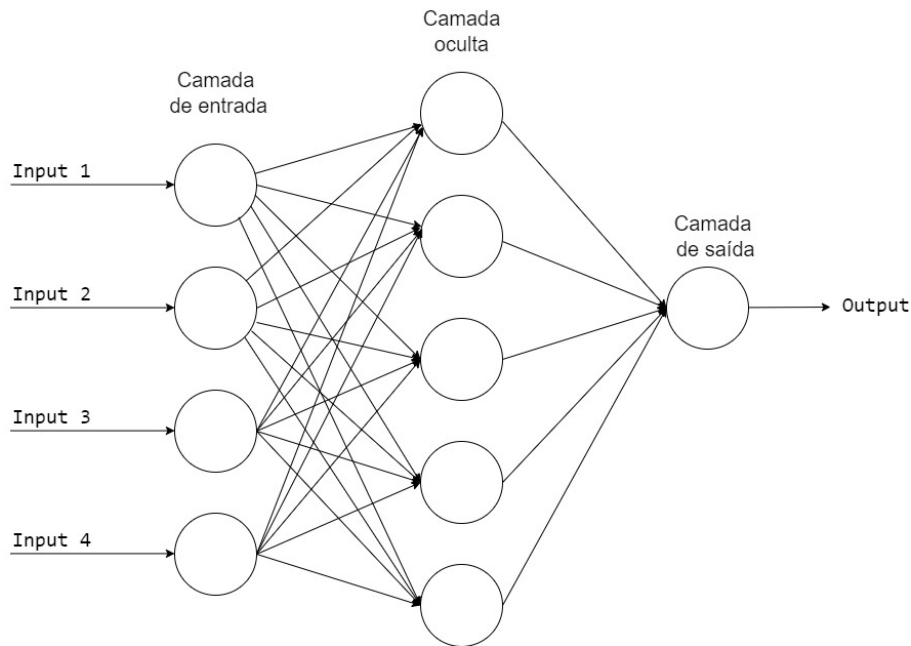


Figura 2.1: Arquitetura da MLP

Cada conexão entre os neurónios é representada por um peso, sendo através que os neurónios comunicam. Esta comunicação apenas ocorre entre neurónios de camadas distintas e nunca entre neurónios da mesma camada. Durante o treino, esses pesos são ajustados para minimizar a diferença entre as previsões da rede e os rótulos reais [Cao et al., 2018].

A seguir são apresentados alguns conceitos relacionados com características dos modelos de ML. Nestes algoritmos existem configurações que podem ser alteradas para controlar e melhorar o seu comportamento, estes são denominados de **hiperparâmetros**. Os hiperparâmetros são parâmetros externos ao modelo de ML que precisam ser definidos antes do treino do modelo. Normalmente são variáveis que definem a estrutura do modelo como, por exemplo, o número de camadas ocultas, função de ativação, função de perda, otimizadores, épocas, entre outros [Silva, 2022].

A **função de ativação** é um componente crucial na construção de uma rede neuronal, pois a sua utilização determina o *output* de um nó que por sua vez servirá de *input* para os próximos nós [Ramachandran et al., 2017]. A utilização desta função permite evitar o envio de informação desnecessária para os próximos neurónios, o que irá impactar o resultado produzido pelo modelo. Esta função é muito utilizada porque introduz não linearidades na rede, permitindo que a rede aprenda padrões mais complexos e não lineares nos dados [Hamdan, 2018]. A relação linear é caracterizada por mudanças consequenciais, ou seja, se uma variável é alterada, a segunda variável sofrerá alterações por consequência. Uma relação é considerada

linear se existe uma proporção constante entre as mudanças nas variáveis [Rasamoelina et al., 2020]. Graficamente, esta relação pode ser representada por uma linha reta num gráfico bidimensional, como é possível observar na Figura 2.2, sendo representada pela Equação 2.1.

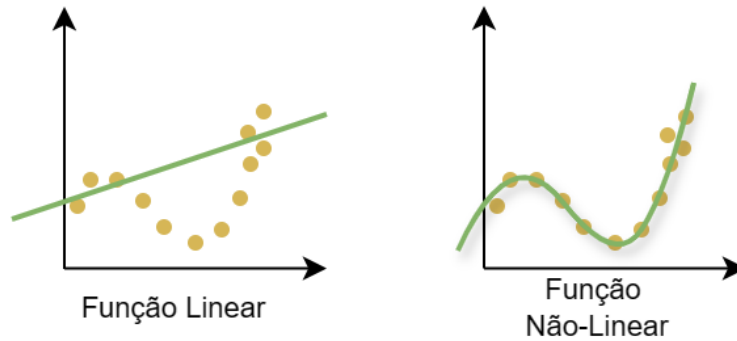


Figura 2.2: Funções linear e não linear

Numa relação não linear, uma alteração na primeira variável pode não impactar a segunda variável. Graficamente, uma relação não linear não pode ser representada por uma linha reta, como é possível verificar na Figura 2.2 sendo representada pela Equação 2.2.

$$\text{Função Linear: } y = mx + b \quad (2.1)$$

$$\text{Função Não Linear: } y = ax^2 + bx + c \quad (2.2)$$

Devido aos avanços na área de ML, existem várias funções de ativação, sendo que a mais utilizada no desenvolvimento de redes neuronais é a função de ativação ReLu. O fator de popularidade deve-se ao facto de ser uma das mais simples [Zhang et al., 2023]. A função ReLU é representada da seguinte forma 2.3:

$$f(x) = \max(0, x) \quad (2.3)$$

A função ReLu é uma função não linear para qualquer valor de entrada x , a saída é $\max(0, x)$, se x for um valor positivo, a saída será x ; se x for um valor negativo a saída será zero. É possível analisar um exemplo de um gráfico da função ReLu na Figura 2.3.

A seleção da função de ativação depende da natureza da tarefa e das características dos dados. São frequentemente identificados critérios para determinar a função de ativação mais adequada que dependem do tipo de rede, do tipo de dados, de entre outras.

A **Função de Perda** (*loss function*) desempenha um papel importante no treino de modelos de ML. É uma métrica que avalia o desempenho do modelo, medindo a

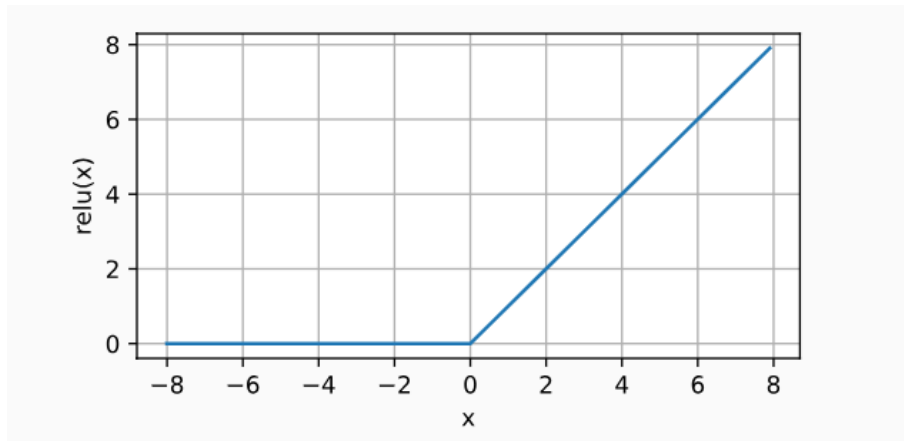


Figura 2.3: Função ReLu [Zhang et al., 2023]

diferença entre as previsões do modelo e os rótulos reais do conjunto de dados. Este valor é necessário para analisar o comportamento do modelo e introduzir melhorias se assim for necessário. A escolha adequada da função de perda é crucial por impactar diretamente a capacidade do modelo de aprender e generalizar a partir dos dados [Wang et al., 2020].

O objetivo fundamental da função de perda é quantificar a discrepância entre as previsões do modelo e os valores reais. Esta discrepância é uma medida da “perda” associada às previsões do modelo. Durante o treino, o objetivo é minimizar essa função de perda, ajustando os parâmetros do modelo para melhor se alinhar aos dados de treino. De entre outras, existem duas funções de perda importantes: a *Mean Squared Error* (MSE) e a *Mean Absolute Error* (MAE) [Wang et al., 2020].

A MSE é apropriada para problemas de regressão e encontra a diferença quadrática média entre o valor previsto e o valor verdadeiro [Han, 2021], como é possível de ver na equação 2.4. Onde N é o número total de observações, o y_i é o valor real (rótulo) da i -ésima observação e o \hat{y}_i é a previsão do modelo para a i -ésima observação.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.4)$$

A MAE é apropriada para tarefas de regressão [Error, 2016]. Baseia-se no cálculo da média das diferenças absolutas entre as previsões (\hat{y}_i) e os rótulos reais (y_i), como é possível de observar na equação 2.5. O MAE é menos sensível a *outliers* em comparação com o MSE.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2.5)$$

O ajuste adequado da função de perda muitas vezes requer vários testes durante o desenvolvimento do modelo. A compreensão da natureza da tarefa, a distribuição dos

dados e as características específicas do problema são fundamentais para a escolha da função de perda mais apropriada. Para problemas de regressão, a MSE é a função mais utilizada.

O processo de otimização tem um grande impacto no processo do treino do modelo de ML, sendo responsável por ajustar parâmetros para minimizar os custos computacionais. Grande parte dos **otimizadores**, utilizam o método clássico que consiste na descida de gradiente, que utiliza o gradiente da função de perda para atualizar iterativamente os parâmetros na direção oposta ao gradiente [Yang and Shami, 2020]. Os otimizadores adaptativos ajustam a taxa de aprendizagem automaticamente com base nas características locais do espaço de parâmetros, acelerando a convergência.

O *Root Mean Square Propagation* (RMSprop) [Kingma and Ba, 2014] é um otimizador muito popular. É um algoritmo de otimização que surgiu para enfrentar desafios associados à escolha adequada da taxa de aprendizagem durante o treino de modelos de ML. Este algoritmo destaca-se pela sua capacidade de realizar uma adaptação dinâmica da taxa de aprendizagem, ajustando-a individualmente para cada parâmetro do modelo. O escalonamento do gradiente é menos agressivo. A soma dos gradientes quadrados é substituída por uma média dos gradientes quadrados. É possível de verificar os cálculos responsáveis através das Equações 2.6 e 2.7.

$$v_{t+1} = \beta v_t + (1 - \beta) g_t^2 \quad (2.6)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_{t+1}} + \epsilon} \cdot g_t \quad (2.7)$$

Nos modelos de ML, o termo **épocas** ou *epoch* refere-se ao número de vezes em que o conjunto de dados vai ser passado pelo algoritmo durante o treino do modelo. Ou seja, uma época é completa quando o algoritmo passa por todas as amostras de treino uma vez. Os modelos são geralmente treinados ao longo de várias épocas para melhorar gradualmente o desempenho, o número ideal de épocas pode variar dependendo do problema, do conjunto de dados e da arquitetura do modelo [Brownlee, 2018]. Treinar por muitas épocas pode levar a *overfitting* (quando o modelo se ajusta demais aos dados de treino e não generaliza bem para novos dados), enquanto treinar por muito poucas épocas pode resultar em subajuste (o modelo não captura adequadamente os padrões nos dados de treino). Portanto, o ajuste do número de épocas é um aspeto importante no treino de modelos de ML.

2.1.2 Avaliação de Modelos

A Avaliação do modelo de ML é uma etapa crucial para determinar a eficácia das suas previsões e a capacidade de generalização para dados não vistos.

Existem vários tipos de métricas de avaliação para cada tipo de problemas. Para problemas de regressão, existem várias métricas de avaliação que podem ser usadas para medir o desempenho do modelo. A utilização do cálculo do MSE e do MAE, já referidas anteriormente, são utilizadas recorrentemente para a avaliação do desempenho do modelo. Mas apesar dessas medidas, existem outras que podem auxiliar na avaliação do desempenho de problemas de regressão como, por exemplo, o Coeficiente de Determinação, representado por R^2 . O R^2 é métrica estatística que avalia o quão bem o modelo de regressão se ajusta aos dados. Um valor de R^2 próximo a 1 sugere que o modelo é uma boa representação dos dados, enquanto um valor próximo a 0 indica que o modelo não está a capturar bem a diversificação da variável dependente [Chicco et al., 2021]. A Equação 2.8 representa o cálculo do R^2 .

$$R^2 = 1 - \frac{\text{Soma dos Quadrados dos Resíduos (SSE)}}{\text{Soma Total dos Quadrados (SST)}} \quad (2.8)$$

Nesta equação, o SSE é a soma dos quadrados dos resíduos, que representa a variabilidade não explicada pelo modelo e o SST é a soma dos quadrados, que representa a variabilidade total na variável dependente. Quando o modelo é perfeito e explica toda a variabilidade, SSE é zero, resultando em $R^2 = 1$. Se o modelo não explica nenhuma variabilidade, SSE é igual a SST, resultando em $R^2 = 0$.

A escolha da métrica dependerá do contexto específico do problema e das preferências em relação à interpretabilidade dos resultados. É comum usar uma combinação dessas métricas para obter uma visão abrangente do desempenho do modelo de regressão.

2.1.3 Problemas comuns associados aos modelos.

No desenvolvimento de modelos de ML, existem problemas que podem ocorrer durante a fase de treino, tais como o *overfitting* e o *underfitting* [Zhang et al., 2019, Schmidhuber, 2015].

O problema de *overfitting* acontece quando o modelo se adaptou muito bem aos dados de treino, mas não generaliza bem para novos dados, ou seja, não consegue lidar com dados que não foram vistos durante o treino. Isso acontece porque o modelo “decorou” o conjunto de dados de treino, mas não aprendeu os padrões importantes para generalizar para novos dados. A complexidade excessiva do modelo, onde existem aglomerados de parâmetros e dados insuficientes são duas condições que podem originar problemas de *overfitting*.

Já o *underfitting* ocorre quando o modelo não se adapta bem sequer aos dados de treino. Isso acontece quando o modelo é muito simples relativamente aos dados que está a tentar modelar e acaba por falhar nessa tarefa. Modelos muito simples, com pouco parâmetros, podem não conseguir representar adequadamente a complexidade dos dados e a falta de treino adequado, ou seja, um número insuficiente de épocas de treino ou uma taxa de aprendizagem muito baixa podem ser causas de problemas de *underfitting*. A Figura 2.4, representa as três situações possíveis.

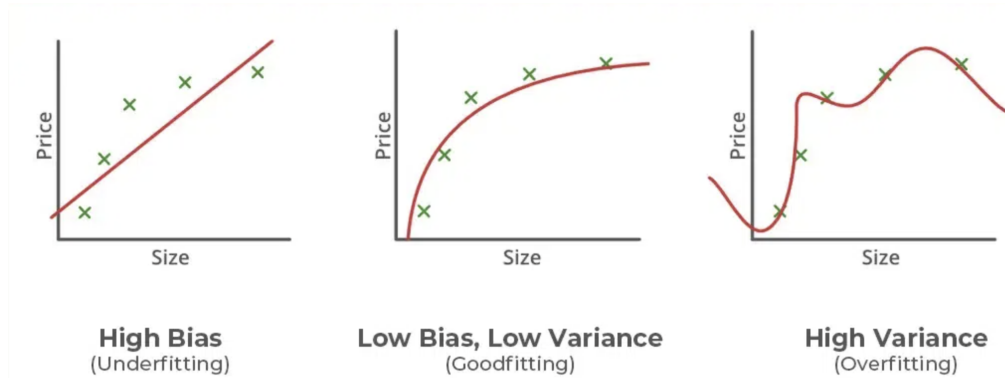


Figura 2.4: Cenário de *underfitting* e *overfitting* [geeksforgeeks, 2023]

Visando combater ou prevenir os erros de *overfitting*, existem várias técnicas. É possível testar diferentes taxas de aprendizagem, quer através da introdução de termos de penalização na função de perda, de forma a limitar os valores dos parâmetros, ou então se for possível aumentar o volume de dados de treino por via de técnicas como rotação, zoom e inversão para proporcionar ao modelo uma visão mais abrangente dos padrões [Jabbar and Khan, 2015].

Para prevenir e corrigir os erros de *underfitting* é sugerido o aumento da complexidade do modelo e melhorar a fase de treino, ajustando o número de épocas de treino e otimizar a taxa de aprendizagem. De forma a desenvolvermos modelos de ML que respondam aos objetivos inicialmente definidos e que obtenham bons resultados finais é crucial encontrar um equilíbrio entre *overfitting* e o *underfitting*.

2.2 Impacto das alterações de peso nas pessoas

Nos últimos anos, o aumento significativo de pessoas com excesso de peso e obesidade tem representado desafios consideráveis para a saúde física, especialmente em relação ao sistema cardiovascular e metabólico [Sun et al., 2022, Cecchini et al., 2010]. Indivíduos com excesso de peso enfrentam frequentemente uma sobrecarga adicional ao coração devido ao aumento do volume sanguíneo e resistência vascular, elevando o risco de hipertensão arterial e doenças cardíacas [Haththotuwa et al., 2020].

O peso, considerado tanto um conceito físico quanto simbólico, desempenha um papel multifacetado na vida das pessoas, influenciando não apenas a saúde física, mas também aspectos psicológicos, sociais e culturais [McKinley, 2017]. No contexto da saúde física, o peso corporal está intimamente relacionado, sendo um fator determinante para o desenvolvimento e prevenção de várias condições médicas [Sung et al., 2019]. O excesso de peso, a obesidade e o baixo peso podem impactar negativamente a saúde, aumentando o risco de doenças cardiovasculares, diabetes, distúrbios metabólicos e outras condições associadas [Santanasto et al., 2017].

Os extremos relativamente ao peso são a obesidade, que se caracteriza pelo excesso de peso, e a diminuição de peso, que pode ser inadequado para o indivíduo. Pessoas com baixo peso frequentemente enfrentam desafios nutricionais que podem comprometer a saúde óssea, aumentando o risco de osteoporose e fraturas. Além disso, a deficiência de nutrientes essenciais pode enfraquecer o sistema imunológico, tornando o organismo mais suscetível a infecções e comprometendo a capacidade de recuperação [Westmoreland et al., 2016].

É crucial reconhecer que a relação entre o peso corporal e a saúde é complexa e individualizada. Fatores genéticos, metabólicos, socioeconômicos desempenham papéis cruciais na determinação da saúde de cada indivíduo. Portanto, uma abordagem personalizada à diversidade de corpos, promovendo uma alimentação equilibrada, a prática regular de atividades físicas e o cuidado com a saúde mental, é essencial [Forbush et al., 2020].

A relação entre o peso e a saúde afeta não apenas a saúde física, mas também a saúde psicológica [Lin et al., 2020]. A pressão social para atender aos padrões estéticos pode ser desafiadora e desmotivadora para quem passa por um processo de perda ou ganho de peso. A luta constante contra estereótipos e normas preconcebidas, cria estigmas adicionais que impactam a saúde mental, influenciando diretamente o bem-estar global. A sociedade contemporânea muitas vezes impõe padrões estéticos que moldam as percepções individuais do corpo. A exposição constante a imagens idealizadas e retocadas cria expectativas muitas vezes inatingíveis relativamente à aparência física [Alimoradi et al., 2020]. A busca incessante por atender a esses padrões pode resultar em pressões psicológicas significativas, levando à insatisfação com a imagem corporal. A comparação constante com padrões irrealistas pode gerar ansiedade, baixa autoestima e até mesmo distúrbios alimentares, impactando diretamente a saúde mental [Rucklidge and Kaplan, 2016, Chi et al., 2021].

Manter o peso corporal estável e saudável é crucial devido ao impacto que esse valor tem na saúde física e psicológica. Para alcançar e manter esse objetivo, é fundamental entender quais fatores podem influenciar as mudanças de peso.

2.2.1 Fatores associados

O peso corporal é influenciado por uma variedade de fatores, incluindo aspectos genéticos, estilo de vida, comportamentos alimentares, atividade física e saúde mental e emocional [Pellegrini et al., 2020]. Os fatores genéticos desempenham um papel crucial na regulação do peso corporal, determinando desde a eficiência do metabolismo energético até a regulação do apetite e da saciedade [Bouchard and Pérusse, 2020, MacLean et al., 2017]. Eles também influenciam a distribuição da gordura corporal, contribuindo para a variação nas formas corporais entre os indivíduos [Rask-Andersen et al., 2019, Hill et al., 2020, Silventoinen and Konttinen, 2020].

Por outro lado, o estilo de vida, incluindo comportamentos alimentares e níveis de atividade física, desempenha um papel significativo nas variações de peso. A disponibilidade de alimentos, a prática de exercícios, o nível de *stress* e a qualidade do sono podem afetar o equilíbrio energético e, conseqüentemente, o peso corporal [Lin et al., 2017, Hayden, 2022, Klaassen et al., 2017]. Atualmente, o estilo de vida predominante é mais sedentário, com uma preferência por refeições rápidas e ricas em gordura saturada, aumentando o risco de doenças crônicas, como obesidade, diabetes e problemas cardiovasculares [Klaassen et al., 2017, Juul et al., 2018].

Além disso, os comportamentos alimentares e os níveis de atividade física são fundamentais para a regulação do peso corporal. Uma dieta equilibrada, composta por macronutrientes e micronutrientes adequados, é essencial para o desenvolvimento físico e mental e a prevenção de doenças crônicas [Di Renzo et al., 2020, Marcason, 2015, Turgeon and Rioux, 2011, Cena and Calder, 2020]. Além disso, a prática regular de atividade física promove o gasto de calorias, o desenvolvimento muscular e a eficiência metabólica [Baguley et al., 2022, Zhao et al., 2024].

A relação entre saúde mental e peso é complexa e bidirecional. Condições como ansiedade, depressão e *stress* podem levar a comportamentos alimentares inadequados, enquanto as mudanças no peso corporal podem afetar a saúde mental. Problemas como o "*emotional eating*" são comuns e podem agravar os desafios relacionados ao peso e à saúde mental [Van Strien, 2018, Frayn and Knäuper, 2018, Konttinen, 2020]. Além disso, o estigma associado ao excesso de peso pode ter impactos negativos tanto na saúde física quanto na mental, aumentando o risco de problemas como depressão, ansiedade e isolamento social [Alimoradi et al., 2020, Lin et al., 2020, Phelan et al., 2015].

É importante reconhecer que dietas rígidas podem levar a problemas psicológicos, como *binge eating*, se não forem acompanhadas por profissionais qualificados. O *binge eating* é comum em pessoas com dietas restritas e pode resultar em sentimentos de perda de controle e compulsão alimentar [de Witt Huberts et al., 2013, Fairburn, 2008].

2.2.2 Acompanhamento das Alterações

O peso, não é apenas um número numa balança, mas sim um indicador de saúde [Conner and Norman, 2017]. A adoção da prática de controlo de peso ajuda na prevenção e identificação precoce de condições de saúde [Rubino et al., 2022]. Essas condições podem ser, síndrome metabólica, distúrbios hormonais ou outras condições médicas subjacentes.

O peso corporal é utilizado na área da nutrição e da saúde como um indicador de saúde física e não só. Ao monitorizar as variações no peso regularmente, é possível obter informação valiosa que vão além da estética e que podem auxiliar na prevenção de doenças associadas a má alimentação [Reinders et al., 2017].

A síndrome metabólica caracteriza-se pela presença de um conjunto de fatores de risco que incluem acumulação de gordura abdominal, tensão alta e níveis elevados de glicose (açúcar) e colesterol no sangue. Em conjunto, estes fatores aumentam muito o risco de doença cardiovascular e diabetes [Carnauba et al., 2017]. A obesidade, por exemplo, está ligada a síndrome metabólica, como resistência à insulina e dislipidemia [Han and Weiss, 2021]. A identificação precoce destes desequilíbrios pode ser feito através do acompanhamento do peso e assim possibilitar intervenções proativas, como modificações na dieta e aumento da atividade física, contribuindo para a prevenção de complicações metabólicas [Verduci et al., 2022].

Os distúrbios hormonais têm impacto no peso e podem ser indicadores de problemas de saúde, tais como hipotireoidismo ou síndrome dos ovários policísticos (SOP) [Barber and Franks, 2021]. O hipotireoidismo acontece devido à tiroide não produzir hormonas em quantidade suficiente, tornando as funções corporais mais lentas. Alguns dos sintomas são cansaço, aumento de peso e sensibilidade ao frio, depressão, pele e cabelo seco e dores musculares [Soares et al., 2020]. A síndrome dos ovários policísticos é um distúrbio hormonal que causa aumento dos níveis de insulina e hormônios masculinos, levando a criação de quistos no ovário e sintomas como menstruação irregular, queda de cabelo, aparecimento de pelos no rosto ou dificuldade para engravidar e por fim aumento de peso [de Souza Pena et al., 2022].

O acompanhamento regular do peso fornece sinais que, quando correlacionados com sintomas específicos, podem apontar para possíveis distúrbios de saúde como, por exemplo, cancro, doenças cardíacas ou distúrbios gastrointestinais [Pearl and Puhl, 2018]. A deteção precoce dessas condições permite uma intervenção adequada e a prevenção de complicações secundárias.

O peso corporal transcende a sua função estética, destacando-se como um indicador valioso da saúde física [Hosseini and Padhy, 2019]. Através do acompanhamento regular, é possível não apenas detetar precocemente desequilíbrios metabólicos, distúrbios hormonais e condições médicas subjacentes, mas também implementar intervenções pro-ativas para prevenir complicações mais sérias [Brannon et al., 2018].

O peso emerge assim como um guardião da saúde física, oferecendo oportunidades contínuas para a promoção da saúde e a prevenção de desafios médicos.

2.2.3 Métricas utilizadas na área da nutrição.

O peso desempenha um papel crucial no nosso bem-estar geral, e a sua importância é reconhecida na procura por um estilo de vida saudável. Para determinar um peso considerado saudável, a medicina e a nutrição desenvolveram várias métricas. Entre elas as ferramentas mais utilizadas pelos profissionais da área são o *Índice de Massa Corporal* (IMC) e a *Taxa Metabólica Basal* (TMB) [Anthanont and Jensen, 2016].

O IMC é uma medida comum na nutrição, utilizada para avaliar o peso relativo de uma pessoa em relação à sua altura. O IMC é calculado dividindo o peso (em quilogramas) pelo quadrado da altura (em metros). A equação para este cálculo é representada pela Equação 2.9.

$$\text{IMC} = \frac{\text{Peso (kg)}}{\text{Altura (m)}^2} \quad (2.9)$$

Após realizar o cálculo, o IMC é interpretado com base nas classificações estabelecidas pela Organização Mundial da Saúde e outras instituições de saúde. Essas classificações dividem o IMC em faixas que indicam diferentes estados de peso. Por exemplo, um IMC abaixo de 18,5 é considerado baixo peso, enquanto valores entre 18,5 e 24,9 indicam peso normal. Pessoas com IMC entre 25 e 29,99 são classificadas como sobrepeso, e aqueles com IMC igual ou superior a 30 são considerados obesos [Khanna et al., 2022].

Embora o IMC seja uma ferramenta útil para avaliar o peso relativo, não é considerado a composição corporal. Isso significa que em alguns casos, indivíduos com alta massa muscular podem ter um IMC mais alto, mesmo sem excesso de gordura corporal. Por esse motivo, é recomendado utilizar o IMC em conjunto com outras medidas, como a avaliação da distribuição de gordura corporal e a análise clínica geral, para uma avaliação mais abrangente da saúde da pessoa.

Outra métrica importante é a TMB, que fornece informações sobre a quantidade mínima de energia necessária para manter as funções vitais do corpo em repouso [Lighton, 2018]. Os nutricionistas utilizam o cálculo da TMB como parte fundamental na definição de planos de nutrição por ajudar a determinar as necessidades calóricas básicas de um indivíduo, tendo em consideração os seus objetivos específicos: manter, perder ou ganhar peso [Liu et al., 2018].

A TMB representa a energia necessária para funções essenciais do organismo, como respiração, circulação sanguínea e regulação térmica, em condições de repouso. Compreender a TMB possibilita a elaboração de recomendações calóricas personalizadas, tendo em conta características individuais como idade, género e peso. A

fórmula mais utilizada para calcular a TMB é a equação de *Harris-Benedict* [Roza and Shizgal, 1984], que tem em consideração o peso, altura, idade e género da pessoa. É importante notar que os valores podem variar dependendo do género do indivíduo, tornando essa equação ainda mais específica para cada caso. A equação de *Harris-Benedict* é expressa pela Equação 2.10.

$$\text{TMB} = \begin{cases} \text{Se for género masculino:} \\ 88.362 + (13.397 \times \text{peso}) + (4.799 \times \text{altura}) - (5.677 \times \text{idade}) \\ \text{Se for género feminino:} \\ 447.593 + (9.247 \times \text{peso}) + (3.098 \times \text{altura}) - (4.330 \times \text{idade}) \end{cases} \quad (2.10)$$

Embora a Equação de *Harris-Benedict* seja amplamente utilizada [Jagim et al., 2018], é importante referir que diferentes versões e adaptações da fórmula existem [Bendavid et al., 2021].

Os métodos nutricionais para avaliação do peso saudável, incluindo a TMB, refletem uma abordagem mais completa e personalizada. Ao considerar variáveis objetivas e subjetivas, nutricionistas podem fornecer recomendações que não apenas promovem um peso corporal saudável, mas também abordam as necessidades energéticas específicas de cada indivíduo, contribuindo para uma prática nutricional mais eficaz e personalizada [Piaggi, 2019].

2.3 Trabalhos Relacionados

Nesta secção vão ser abordadas algumas soluções já existentes de modelos de previsão na área da saúde e que recorram a técnicas de ML. O número de estudos relacionados à tarefa de classificação de ganho de peso é relativamente limitado. A aplicação de modelos ML no campo da saúde representa um avanço significativo na análise de dados médicos. Essa abordagem permite a deteção precoce de condições médicas, a personalização de tratamentos com base em perfis genéticos individuais e a otimização de processos hospitalares. Por meio de algoritmos de ML, os profissionais de saúde podem processar grandes conjuntos de dados clínicos, identificar padrões úteis e fornecer diagnósticos mais precisos. Além disso, a análise preditiva contribui para a prevenção de surtos epidemiológicos e de doenças. A sinergia entre a saúde e a utilização de modelos de ML promove uma abordagem mais pro-ativa e personalizada na medicina, melhorando significativamente a qualidade dos cuidados prestados aos pacientes.

Em [Sonawane and Patil, 2014] pretendeu-se prever o risco de doenças cardíaca utilizando técnicas de MLP. A rede neural neste sistema aceita 13 características clínicas como entrada, sendo treinada utilizando o algoritmo *back-propagation* [Li

et al., 2012] para prever a presença ou a ausência de doenças cardíacas no paciente. O *dataset* utilizado para alimentar este modelo, é constituída por 303 registos, cada um com 13 atributos clínicos que incluem idade, sexo, tipo de dor no peito, pressão arterial em repouso, colesterol, açúcar no sangue em jejum, ECG em repouso e frequência cardíaca máxima. Neste conjunto de dados, 164 pertencem à categoria saudável e 139 pertencem à categoria de doença cardíaca. Os resultados obtidos para prever a presença ou ausência de doença cardíaca nos pacientes têm uma precisão máxima de 89%, o que não só ajudou os médicos a prever e a diagnosticar com precisão várias doenças, mas também ajuda os pacientes com um diagnóstico precoce. Este trabalho propôs um modelo fiável que pode garantir a segurança e a saúde de um indivíduo.

No âmbito da previsão de riscos de obesidade, em [Ferdowsy et al., 2021] foi utilizada também uma abordagem baseada em modelos de ML. Foram recolhidos mais de 1100 dados de pessoas de diferentes idades, tanto de pessoas que sofrem de obesidade como de pessoas que não sofrem de obesidade. Para o desenvolvimento deste projeto, foram aplicados nove algoritmos de ML, *k-nearest neighbor*, *Random Forest* (RF), *Logistic Regression* (LR), MLP, *support vector machine*, *naïve Bayes*, *adaptive boosting*, *decision tree* e *gradient boosting classifier*. Para cada um destes algoritmos foram aplicadas métricas de desempenho. A partir dos resultados experimentais, foram determinados 3 parâmetros para classificar a obesidade: alta, média e baixa. Em termos comparativos, o algoritmo de LR atinge a precisão de 97,09%, sendo a maior de entre todos os classificadores.

Num estudo conduzido por [Iwendi et al., 2020] foi proposta a implementação de uma solução de *Deep Learning* (DL) para um conjunto de dados de saúde. Este sistema visa identificar automaticamente a dieta mais adequada para cada paciente, baseando-se na sua condição de saúde e em outras características individuais, como idade, sexo, peso e composição nutricional dos alimentos. A investigação concentrou-se na aplicação de algoritmos de DL e ML, incluindo regressão logística, *Naive-Bayes*, RNN, MLP, e *Long Short-Term Memory* (LSTM). Estes algoritmos foram aplicados a um conjunto de dados médicos compostos por informações de 30 pacientes com 13 características distintas de diferentes doenças e 1000 produtos alimentares. Os resultados dos testes e treinos dos vários modelos demonstraram a eficácia das técnicas de DL e ML nesta aplicação específica. O modelo LSTM destacou-se em termos de precisão, *recall* e medida F1, alcançando 98% de precisão, 99% de *recall* e 80% de medida F1 para a classe permitida, e 89% de precisão, 73% de *recall* e 80% de medida F1 para a classe não permitida.

Em [Zhao et al., 2021] é proposta a avaliação e a comparação do desempenho de quatro algoritmos diferentes de ML na previsão do risco de desenvolver hipertensão. Esta previsão será realizada a partir de informações fáceis de reunir. O objetivo final é escolher o melhor algoritmo de ML para desenvolver um modelo de previsão

de risco de hipertensão. Os quatro algoritmos de ML utilizados neste estudo foram RF, CatBoost, redes neurais, MLP e LR. O *dataset* utilizado para construir o modelo deste projeto é proveniente de um centro de exames físicos de um hospital em Pequim, na China. Foram coletados 29.750 casos de dados completos. Entre eles, há 10.650 casos de hipertensão e 19.100 casos normais. A maioria dos casos normais tem entre 18 e 70 anos, e a maioria dos casos de hipertensão tem entre 20 e 75 anos. Através da comparação de várias variáveis e aplicações de métricas de avaliação foi possível de verificar que aplicação do modelo de RF superou os modelos restantes. Os resultados revelaram que o modelo de RF conseguia distinguir populações de hipertensão de alto e baixo risco com base em variáveis fáceis de reunir, tornando-se assim o modelo que tem mais valor na aplicação controlo da hipertensão.

Em [Nagari and Inayati, 2020] foi aplicado um algoritmo de *K-means* para determinar os valores nutricionais de crianças na Indonésia, com o objetivo de agrupar um conjunto de crianças que necessitam de maior apoio. Os resultados deste estudo indicam que o estado nutricional destas crianças pode ser agrupado em 4 *clusters* com base no peso relativamente à idade. O *cluster* 1 inclui 23 crianças em situação de desnutrição, o *cluster* 2 conta com 17 crianças em mau estado nutricional, o *cluster* 3 com 7 crianças também em mau estado nutricional, e o *cluster* 4 com 10 crianças.

[Rostami et al., 2022] desenvolveram um sistema híbrido de recomendação de alimentos, dividido em duas fases: uma baseada no conteúdo alimentar e outra no utilizador. Na primeira fase, utilizaram agrupamento de gráficos, enquanto na segunda fase adotaram uma abordagem baseada em DL para agrupar utilizadores e alimentos. Os resultados mostraram que o sistema híbrido é, em média, 11,8% mais eficaz do que o método baseado apenas no utilizador e 11,9% mais eficiente do que o método baseado apenas no conteúdo alimentar. Além disso, ao incluir a rede de confiança do utilizador, houve um aumento de desempenho médio de 5,3%. Testes também indicaram que considerar o fator tempo resultou num desempenho global 9,51% superior. As métricas de avaliação do modelo proposto foram significativamente melhores do que as dos modelos de comparação.

Em [Raut et al., 2018] é proposto um sistema que utiliza raciocínio baseado em regras para eficientemente fazer recomendações alimentares e nutricionais, enquanto emprega um algoritmo genético para a geração de menus de refeições [Karaboga, 2010]. Este sistema depende da API do Google FIT do utilizador para obter informações sobre as atividades diárias e as necessidades energéticas. Além disso, baseia-se no histórico de doenças anteriores do paciente para fazer recomendações personalizadas de dieta. Com base nos resultados obtidos, os autores concluíram que os algoritmos de DL apresentaram um melhor desempenho em comparação com os algoritmos de ML.

[Weber and Achananuparp, 2016] realizaram um estudo utilizando dados alimentares diários de mais de 4.000 utilizadores ativos a longo prazo da aplicação MyFitnessPal, visando entender as características de dietas mal sucedidas e apoiar pessoas que buscam perder peso e manter a saúde. Foi treinado um modelo de ML para prever o excesso ou falta de calorias diárias definidas pelos utilizadores, observando os recursos que contribuem para essa previsão. O estudo focou num classificador que prevê se os utilizadores ultrapassarão ou ficarão abaixo da sua meta de calorias, com base nos alimentos consumidos. Os resultados demonstraram a viabilidade de utilizar esses dados para análises relacionadas com a saúde, indicando que dietas personalizadas e adaptativas, geradas automaticamente com base em padrões de estilo de vida e atividade dos utilizadores, têm potencial promissor.

No contexto do desenvolvimento de aplicações, [Leipold et al., 2018] criaram um sistema de suporte nutricional que analisa os hábitos alimentares do paciente e se adapta às mudanças comportamentais utilizando diversas abordagens persuasivas, como auto-monitorização, personalização, reflexão, recomendações e rastreamento. Embora os sistemas automatizados de recomendação alimentar possam oferecer benefícios significativos em comparação com os nutricionistas humanos, enfrentam desafios como usabilidade, eficiência, eficácia e satisfação do utilizador. O estudo destacou a necessidade de melhorias em vários aspetos da aplicação, incluindo recomendações, desempenho e facilidade de rastreamento da ingestão de alimentos. Apesar das limitações, 43% dos participantes demonstraram interesse em utilizar regularmente a aplicação, sendo que a maioria (85%) preferiu o rastreamento alimentar diário em vez de um questionário alimentar semanal. Os autores ressaltaram a importância de integrar informações contextuais e sociais, além de melhorar a precisão dos dados de entrada recebidos.

Uma outra aplicação, denominada DIETOS foi desenvolvida por [Agapito et al., 2016]. É uma aplicação que oferece suporte à organização dietética. Utilizando uma pesquisa dinâmica elaborada por médicos e compilada pelos utilizadores, o sistema cria perfis individuais. Além de recomendar alimentos similares e saudáveis, também fornece orientações nutricionais para diversos problemas de saúde. Todos os dados são acessíveis por meio de uma interface web intuitiva, onde os alimentos compatíveis são destacados em verde e os incompatíveis em vermelho. A obtenção de informações sobre a produção dos alimentos é facilitada por interação gráfica num mapa. Os resultados, incluindo alimentos típicos da localização selecionada, são exibidos no mapa, permitindo acesso às propriedades nutricionais de cada produto.

2.4 Conclusão

Neste capítulo foram expostas informações acerca da importância do peso na vida das pessoas, através dos fatores associados às alterações de peso, o impacto

que este fator tem na saúde humana e o modo como essas alterações podem ser acompanhadas. Foram também abordados os métodos de cálculo do peso ideal, através da exposição do IMC e da TMB. Estas duas métricas possibilitam auxiliar o profissional da área da saúde e da nutrição no acompanhamento de pacientes. Tendo em conta que a aplicação de modelos de ML na área da saúde representa um avanço na análise de dados médicos, foram também abordados os fundamentos teóricos relacionados com o ML. Esses fundamentos incluíram as configurações que podem ser alteradas para controlar e melhorar o comportamento dos algoritmos, nomeadamente os hiperparâmetros. Também foram abordadas as funções de ativação, funções de perda, otimizadores e épocas. Uma vez que a avaliação do modelo de ML é uma etapa fundamental para determinar a eficácia das previsões do modelo, foram expostos os diversos tipos de métricas de avaliação para cada tipo de problema. Por fim foram analisados vários trabalhos relacionados a área da saúde, da nutrição e da aplicação de modelos e técnicas de ML. Tendo em conta os fundamentos teóricos abordados e os trabalhos relacionados enquadrados no tema em análise, serão aplicados modelos de redes neuronais MLP com ênfase na área da previsão de peso tendo em conta o estilo de vida de um indivíduo.

Capítulo 3

Metodologia

O presente capítulo aborda a metodologia adotada para a base do desenvolvimento do modelo de ML. A Criação de modelos de ML obedece a uma metodologia que permite descrever o processo seguido para a criação do modelo e a respetiva avaliação.

Foi utilizada a metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [Chapman et al., 1999] para a realização deste projeto. Esta metodologia foi selecionada por ser adaptativa uma vez que permite a iteração dinâmica entre fases. Este aspeto permite retornar às fases anteriores para ajustar e melhorar a abordagem com base nos resultados ao longo do processo [Purbasari et al., 2021]. Permite, também, um ajuste da metodologia conforme necessário para se adequar ao contexto específico do projeto. A análise constante em cada fase à medida que os resultados preliminares são obtidos, permite ajustar as abordagens utilizadas, garantindo que o projeto permaneça alinhado com os objetivos do negócio. Este processo iterativo pode ser observado na Figura 3.1.

A metodologia CRISP-DM assenta em 6 fases: compreensão do negócio, compreensão dos dados, preparação dos dados, modelação, avaliação e implementação [Schröer et al., 2021]. A última etapa consiste na fase de instalação no ambiente de produção e para o desenvolvimento deste projeto esta etapa não se realizou.

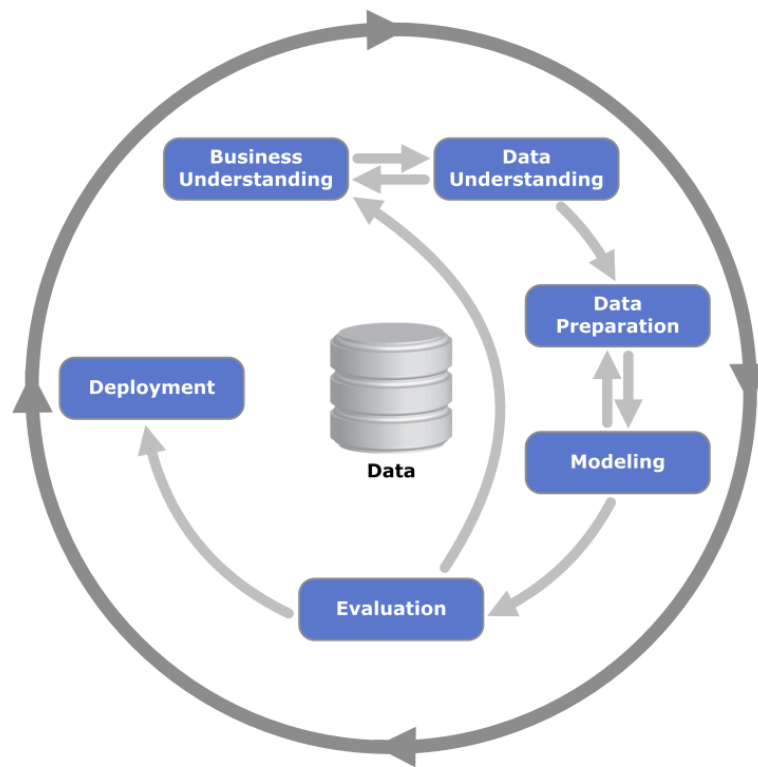


Figura 3.1: Diagrama CRISP-DM Fonte:[Chapman et al., 2000]

3.1 Compreensão do negócio

Nesta etapa vão ser compreendidos o contexto e os objetivos do negócio, definindo quais os objetivos e necessidades do projeto. Este processo inicia-se com um estudo sobre a área de negócio onde o modelo vai atuar. Este projeto atuará na área da saúde e visa compreender o impacto que o estilo de vida do ser humano tem no seu peso corporal.

O peso corporal é um indicador crucial da saúde e do bem-estar, sendo fundamental entender os fatores que contribuem para a sua variação de modo a desenvolver abordagens eficazes de gestão de peso [Bellicha et al., 2021]. A saúde alimentar é um fator importante para garantir uma boa qualidade de vida, uma vez que fornece energia e bem-estar geral [Alves and de Oliveira Cunha, 2020]. Com a adoção de uma alimentação saudável, aliada à prática de exercício físico, é possível: manter um peso corporal saudável, ter um bom desenvolvimento físico e mental e prevenir doenças.[Cena and Calder, 2020].

Assim sendo, com o desenvolvimento deste projeto pretende-se alertar para as consequências que os maus hábitos alimentares podem ter no peso futuro e fazer uma avaliação da sua rotina com base na sua alimentação e do exercício físico.

3.2 Compreensão dos dados

No que diz respeito à etapa de compreensão dos dados, esta contempla a pesquisa nos repositórios de dados existentes, de modo a entender os dados disponíveis e necessários para responder aos objetivos do projeto. Durante a fase de compreensão dos dados, os mesmos são analisados tendo em conta os objetivos do trabalho e a compreensão do negócio.

A etapa de compreensão dos dados, contempla a pesquisa no *dataset* por atributos que respondam às necessidades do projeto. Para tal aqui serão analisadas as seguintes características dos dados: como o formato, tamanho, tipos de variáveis. Esta análise permite avaliar a qualidade dos dados, sendo possível também extrair as primeiras impressões sobre o conjunto de dados e detetar padrões.

3.3 Preparação dos dados

A fase de preparação de dados é composta por quatro etapas que servem para orientar a construção do *dataset* final que irá alimentar as fases seguintes do trabalho. Estas fases estão ilustradas na Figura 3.2, sendo elas: a limpeza dos dados, transformação dos dados, engenharia de recursos e seleção de recursos, tal como é ilustrado.

O processo de limpeza de dados é necessário para remover inconsistências nos dados. Esse processo envolve o tratamento de valores ausentes, duplicados e *outliers*. Seguidamente, realiza-se a transformação dos dados para que estes fiquem no formato correto para a modelação. Nesta fase é possível converter variáveis categóricas em formatos adequados para o algoritmo de ML, recorrendo a técnicas de normalização e padronização.

A Engenharia de atributos consiste na seleção, criação ou conversão de atributos que possam contribuir para a precisão do modelo. Para isso é possível proceder à criação de novas variáveis a partir das existentes no *dataset*. Por sua vez, a seleção de atributos, identifica as variáveis mais relevantes para a criação do modelo.

Será nesta etapa que o conjunto de dados irá ser dividido em dois conjuntos, um de teste e outro de treino. Estes dois conjuntos servem para treinar e testar o modelo, de forma a obtermos os resultados.

As etapas descritas pelo CRISP-DM fazem parte de um processo iterativo, onde o desempenho do modelo pode contribuir para ajustes ao nível da estrutura dos atributos.

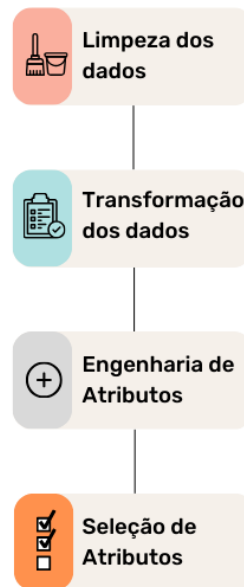


Figura 3.2: Diagrama do processo de preparação dos dados.

3.4 Modelação

Nesta fase é criado um modelo de regressão que permita prever o peso de uma pessoa tendo como base: o seu histórico alimentar, a sua prática de exercício físico e a sua ingestão diária de alimentos no dia atual. Esta etapa está muitas vezes em *loop* com a preparação dos dados, até que seja obtido um modelo funcional.

A etapa de modelação começa com a escolha do algoritmo mais adequado ao problema em questão. Passando para a configuração e parametrização do modelo. Nesta etapa é necessário configurar os hiperparâmetros que podem afetar significativamente o desempenho do modelo. Com o algoritmo escolhido e os parâmetros configurados, é hora de criar a estrutura do modelo. Isso envolve a definição da arquitetura do modelo, incluindo o número de variáveis de entrada, camadas ocultas e a camada de saída. Uma vez que a estrutura do modelo esteja definida, o modelo é treinado usando os dados disponíveis. Durante o treino, o modelo ajusta os seus parâmetros para minimizar uma função de perda. Após o treino, o modelo é testado usando dados de teste para avaliar o seu desempenho e capacidade de generalização.

Em suma, o objetivo da etapa de modelação é criar e validar o modelo propostos para responder ao objetivo do projeto.

3.5 Avaliação

Após terminada a fase de construção do modelo, segue-se a avaliação do desempenho do modelo. Para tal, é necessário testar e analisar os resultados, para perceber se os mesmos correspondem ao objetivo do projeto e se é necessário proceder à revisão do processo. Esta avaliação será realizada recorrendo a métricas de avaliação utilizadas em modelos de regressão, como MAE, MSE e o coeficiente de determinação (R^2). É importante, também, analisar os erros cometidos pelo modelo para identificar padrões ou tendências. Isso pode ajudar a entender as limitações do modelo e fornecer *insights* sobre as áreas onde pode ser melhorado.

Caso algum objetivo não esteja totalmente concluído, o modelo permite sempre voltar ao início do processo. No final desta fase é expetável que seja tomada a decisão, caso os resultados do modelo estejam prontos para ser aplicados.

3.6 Considerações Finais

Este capítulo descreve todas a etapas da metodologia CRISP-DM, para o desenvolvimento do modelo. Iniciando-se com a justificação da escolha desta metodologia e descrição de cada etapa. Na Secção 3.1 foi explicado a importância da compreensão de negócio. Nas secções seguintes foram abordadas as técnicas de compreensão de dados, seleção de dados, modelação e avaliação respetivamente.

No próximo capítulo será explicado detalhadamente como cada etapa foi executada no desenvolvimento do modelo.

Capítulo 4

Implementação

O presente capítulo tem em vista apresentar a implementação do modelo de ML e a respetiva avaliação do modelo proposto neste estudo. Utilizando a metodologia CRISP-DM como guia para o desenvolvimento do modelo ML, neste sentido, o capítulo encontra-se estruturado conforme as etapas da metodologia apresentada. Começando pela compreensão dos dados, a seguir a preparação dos dados, modelação e por fim a avaliação.

4.1 Compreensão dos dados

Este projeto envolve a utilização de um *dataset* com dados associados ao registo diário do peso corporal, de calorias consumidas e gastas por um indivíduo durante um determinado período.

O *dataset* utilizado para a criação do modelo é intitulado de “*2018 calorie, exercise and weight changes*”, estando publicado no repositório público Kaggle [CHRISBOW, 2018]. O *dataset* é composto por 14 colunas, que estão representadas na Tabela 4.1. Os dados foram extraídos de um ficheiro em formato *Common Separated Values* (CSV), contendo 152 registos e 14 variáveis.

Este *dataset* corresponde ao registo de uma pessoa que iniciou uma dieta para perder peso no início de janeiro, após o período do Natal. Este registo possui as seguintes informações: peso diário, a ingestão aproximada de calorias, e as calorias gastas, alguma informação sobre os alimentos que ingeria como, por exemplo, se

comera cinco donuts, se bebera um copo de vinho ou se tinha feito uma alimentação rica em proteína, se fez uma corrida e o tipo de exercício físico que praticava.

A análise das características do conjunto de dados foi realizada através da descrição fornecida pelo autor. Nessa descrição é possível ter conhecimento sobre as características dos dados, como o seu formato, tamanho, tipos de variáveis, distribuição e valores em falta.

Tabela 4.1: Definição dos atributos do *dataset*

Date	date of observations
Stone, Pounds, Ounces	Stone, pounds, ounces: observed weight
Weight_oz	Total weight in ounces
Calories	Approximate calories consumed on that date. Calorie counts were taken for each food based on nutritional information
cals_per_oz	calories per ounce - calories for that day divided by weight in ounces
five_donuts	day where main meal was bag of five jam donuts, 1 = TRUE, 0 = FALSE
walk	day included at least one brisk walk of over 20 minutes, 1 = TRUE, 0 = FALSE
run	day included at one short run of 2.5 miles, 1 = TRUE, 0 = FALSE
wine	day finished with at least one large glass of wine when not accompanied by other fluids, 1 = TRUE, 0 = FALSE
prot	day involved what would be considered a high protein diet, 1 = TRUE, 0 = FALSE
weight	signifies day during period of weight-based exercise; includes rest days as part of period, 1 = TRUE, 0 = FALSE
change	change: resulting change in weight in oz, calculated from following day's observed weight

4.2 Preparação dos dados

A preparação de dados é a etapa onde foram reunidos e transformados os dados necessários para o desenvolvimento do modelo. Os dados escolhidos respondem às necessidades do modelo na medida em que disponibilizam o registo do número de calorias diário, do peso corporal e as alterações do mesmo ao longo do tempo. Após uma análise ao *dataset*, foram escolhidas 3 colunas das 14 apresentadas pelo autor.

A Tabela 4.2 apresenta as colunas selecionadas para o desenvolvimento do projeto, onde é explicado o conteúdo de cada variável.

Tabela 4.2: Detalhes dos Dados escolhidos

weight oz	Total weight in ounces
calories	Approximate calories consumed on that date Calorie counts were taken for each food based on nutritional information
change	Resulting change in weight in oz, calculated from following day's observed weight

A preparação e análise dos dados foi realizada com recurso à linguagem de programação Python, utilizando a biblioteca *Pandas*. Esta ferramenta foi utilizada no desenvolvimento deste projeto para a leitura do *dataset* original e para a manipulação e organização dos dados e para criar um *dataframe* com as variáveis desejadas. A Figura 4.1 representa os dados após o seu processamento.

Devido à impureza dos dados foi necessário proceder a uma limpeza e normalização dos mesmos, uma vez que o ficheiro original continha irregularidades, por exemplo, existiam valores alfanuméricos em colunas numéricas. Após este processo o *dataset* ficou reduzido a 139 registos.

Index	CaloriasIngeridas	CaloriasGastas	Peso
0	1950	-30	2726
1	2600	8	2696
2	2500	0	2704
3	1850	-40	2704
4	2900	14	2664
5	3600	14	2678
6	2400	-2	2692
7	3100	6	2690
8	2200	-8	2696
9	1800	-40	2688
10	2300	-18	2648
11	3000	12	2630
12	4000	34	2642
13	2800	-12	2676

Figura 4.1: Tabela com os dados após o pré-processamento

Com o objetivo de calcular a taxa metabólica basal para cada registo do *dataset*, foi necessário recorrer à função de *Harris and Benedict*, representada pela Equação 2.10, sendo adicionada posteriormente uma nova coluna com esse cálculo. Esta função usa a altura de uma pessoa, o seu peso, idade e género, para calcular a sua taxa metabólica basal.

No desenvolvimento deste projeto foi apenas utilizado o cálculo referente ao género masculino, pois não foram encontrados *datasets* realizados por indivíduos do género feminino.

Para proceder ao cálculo desta fórmula eram necessários os dados referentes à idade (*age*), altura (*height*) e peso (*weight*).

Após reunidas todas as condições para o cálculo da função de *Harris and Benedict*, foi criada uma coluna no *dataset*, designada de 'TMB' que armazena o resultado do cálculo da taxa metabólica basal para cada linha do *dataset*. O cálculo da taxa metabólica basal para cada registo está representado na Figura 4.2.

Index	CaloriasIngeridas	CaloriasGastas	Peso	TMB
0	1950	-30	2726	37340.1
1	2500	8	2696	36938.2
2	2500	0	2704	37045.3
3	1850	-40	2704	37045.3
4	2900	14	2664	36509.5
5	3600	14	2678	36697
6	2400	-2	2692	36884.6
7	3100	6	2690	36857.8
8	2200	-8	2696	36938.2
9	1800	-40	2688	36831
10	2300	-18	2648	36295.1
11	3000	12	2630	36054
12	4000	34	2642	36214.7
13	2800	-12	2676	36670.2

Figura 4.2: Tabela com o cálculo da taxa metabólica basal (TMB)

No final de todos os dados reunidos, foi criada uma função para calcular o peso futuro tendo em conta os resultados do peso, da taxa metabólica basal e das calorias ingeridas e gastas nos últimos três dias. Esta função encontra-se estruturada nas seguintes fases:

- criação de colunas com os valores temporalmente desfasados: a partir do registo atual das calorias gastas, calorias ingeridas, do peso e da taxa metabólica basal, vão ser criadas colunas que representam os valores das colunas existentes nos dias $t-n$, sendo t a referência temporal do registo atual e n o número de dias. É possível verificar estas colunas na Figura 4.3
- cálculo do indicador 'FutureWeight': é introduzida uma nova coluna chamada 'FutureWeight', que representa o cálculo da previsão do peso futuro tendo em conta os valores das colunas $t-1, t-2$ e $t-3$.
- limpeza e preparação do conjunto de dados: são realizados ajustes para garantir que o novo conjunto de dados seja coerente, incluindo a remoção de linhas com dados ausentes.
- resultado final: o resultado final é um conjunto de dados maior e mais detalhado, que facilita a compreensão das mudanças de peso ao longo do tempo. Este novo conjunto de dados está apresentado na Figura 4.3.

Index	CaloriasIngeridas-t1	CaloriasGastas-t1	Peso-t1	TMB-t1	CaloriasIngeridas-t2	CaloriasGastas-t2	Peso-t2	TMB-t2	CaloriasIngeridas-t3	CaloriasGastas-t3	Peso-t3	TMB-t3	FutureWeight
0	2600	8	2696	36938.2	2580	0	2784	37045.3	1850	-40	2784	37045.3	2786.05
1	3500	0	2784	37845.3	1850	-40	2784	37045.3	2900	14	2664	36589.5	2675.21
2	1850	-40	2784	37845.3	2900	14	2664	36589.5	3600	14	2678	36697	2680.11
3	3900	14	2664	36589.5	3600	-2	2678	36697	2400	-2	2692	36884.6	2678.65
4	3600	14	2678	36697	2400	-2	2692	36884.6	3100	6	2690	36857.8	2638.05
5	3400	-2	2692	36884.6	3100	6	2690	36857.8	2200	-8	2696	36938.2	2655.99
6	3100	6	2690	36857.8	2200	-8	2696	36938.2	1800	-40	2688	36831	2671.59
7	3200	-8	2696	36938.2	1800	-40	2688	36831	2300	-18	2648	36295.1	2671.81
8	3000	-40	2688	36831	2300	-18	2648	36295.1	3000	12	2630	36054	2675.58
9	2300	-18	2648	36295.1	3000	12	2630	36054	4800	34	2642	36214.7	2661.51
10	3000	12	2630	36054	4800	34	2642	36214.7	2800	-12	2676	36670.2	2620.1
11	4000	34	2642	36214.7	2800	-12	2676	36670.2	2300	-2	2664	36589.5	2684.06
12	3000	-12	2676	36670.2	2300	-2	2664	36589.5	2400	-4	2662	36482.7	2620.52
13	2300	-2	2664	36589.5	2400	-4	2662	36482.7	2100	-10	2658	36429.1	2656.53
14	2400	-4	2662	36482.7	2100	-10	2658	36429.1	2450	10	2648	36295.1	2644.13
15	3100	-10	2658	36429.1	2450	10	2648	36295.1	3400	22	2658	36429.1	2639.35
16	3450	10	2648	36295.1	3400	22	2658	36429.1	3600	14	2680	36723.8	2631.13
17	3400	22	2658	36429.1	3600	14	2680	36723.8	4200	16	2694	36911.4	2616.15
18	3600	14	2680	36723.8	4200	16	2694	36911.4	2000	-18	2710	37125.7	2630.03
19	4200	16	2694	36911.4	2000	-18	2710	37125.7	2150	-20	2692	36884.6	2656.08
20	3900	-18	2710	37125.7	2150	-20	2692	36884.6	2300	-16	2672	36616.6	2675.39

Figura 4.3: Tabela com o cálculo do peso futuro tendo em conta os últimos três dias

No final deste processo obtém-se o *dataset* final pronto, sendo assim possível avançar para a fase da criação de dois conjuntos de dados, que vão ser utilizados para treino e teste do modelo.

Esses dois conjuntos, são designados de *train split* e um *test split*. O *train split* permite treinar o modelo e corresponde a 70% do volume dos dados, enquanto o *test split* representa os restantes 30%, para prever o *target* e avaliar a desempenho do modelo.

4.3 Modelação

Quando é terminada a fase de compreensão e preparação de dados é possível avançar para a etapa de desenvolvimento do modelo. O modelo deve calcular o peso futuro tendo em conta o peso do indivíduo, as calorias ingeridas e gastas nos últimos três dias.

A arquitetura da rede neuronal está dividida em três grupos: uma camada de entrada, a(s) camada(s) intermédia(s) e uma camada de saída. A arquitetura do modelo está representada na Figura 4.4.

A camada de entrada, é composta por vinte neurónios e utiliza a função *ReLU* como função de ativação. É através desta camada que entra a informação na rede que depois será distribuída para as camadas intermédias, através de uma ligação representada por pesos.

A camada interna é composta por cinco neurónios, tendo sido escolhida a função *ReLU* como função de ativação para introduzir não linearidades e permitir que o modelo aprenda representações mais complexas. É nesta camada que ocorre o processo para a resolução dos problemas.

Na camada de saída, foi utilizado apenas um neurónio, sendo o resultado da previsão do peso futuro e como é problema de regressão, onde a saída desejada é

um valor contínuo. Não é aplicada uma função de ativação específica na camada de saída, ao ser necessária uma saída contínua. Na Tabela 4.3 são apresentados os hiperparâmetros utilizados.

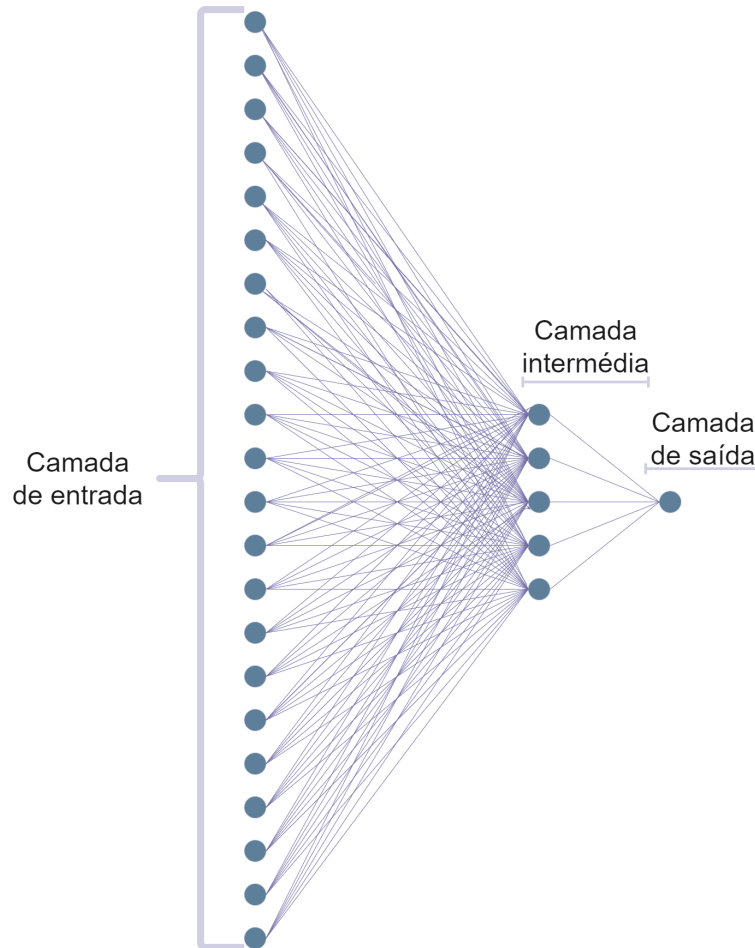


Figura 4.4: Diagrama representativo do modelo de rede neuronal implementado

Hiperparâmetros	Valores Utilizados
Épocas	2000
Nº de neurónios na camada de entrada	20
Nº de neurónios na camada interna	5
Nº de neurónios na camada saída	1
Função de Ativação	ReLU
Função de Perda	MSE, MAE
Otimizadores	RMSprop

Tabela 4.3: Valores dos Hiperparâmetros.

O processo de treino do modelo é inicializado com o valor das variáveis de entrada a serem inseridos na rede neuronal, onde vão ser treinados. Durante este processo os dados de teste são utilizados para que o modelo possa fazer verificações daquilo que está a aprender. Aqui é possível obter o resultado de saída, ou seja, a previsão do peso futuro que fica armazenado na variável Y_pred .

Os resultados dessa variável vão ser avaliados em comparação com o Y_test , que correspondem ao conjunto de dados definidos para teste. Para avaliar se o modelo aprendeu corretamente são utilizadas métricas de avaliação como o MSE e o MAE. Este processo pode ser analisado na Figura 4.5.

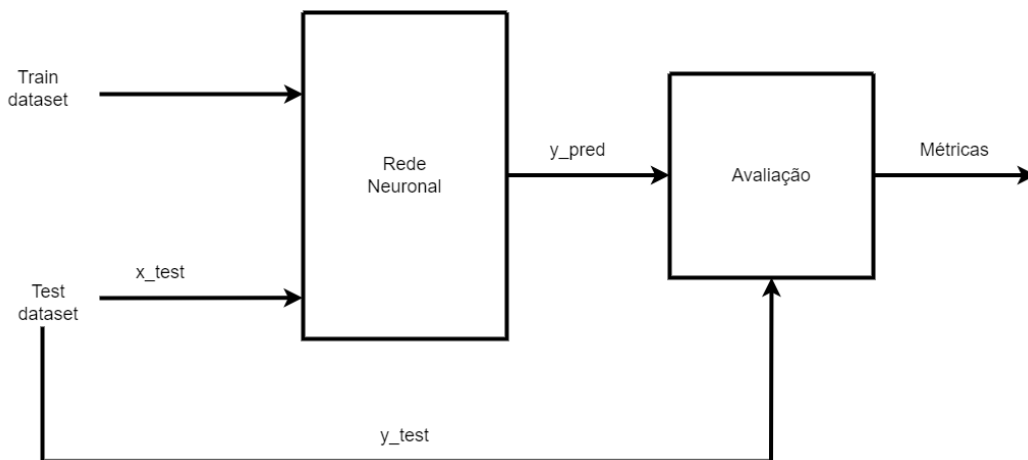


Figura 4.5: Diagrama do processo de treino e de teste do modelo

4.4 Avaliação

Para avaliar o correto desempenho do modelo são utilizadas várias métricas de avaliação, nomeadamente, o MAE, o MSE, Coeficiente de Determinação (R^2) e a comparação dos dados previstos e dos dados reais.

A avaliação foi realizada em termos da capacidade de generalização do modelo e do erro de previsão.

4.5 Generalização do modelo

Para a avaliação foi utilizada a função de *Loss*, para medir o desempenho do modelo a cada iteração, uma vez que traduz o quanto o resultado previsto do modelo difere do resultado real. Durante a fase de treino, a função de *Loss* recorre ao cálculo do MAE. Depois da fase de treino, é calculado o MSE.

Analisando os dois gráficos presentes na Figura 4.6 é possível verificar que o erro de generalização dos modelos é reduzido. Ou seja, o modelo tem um comportamento semelhantes com os dados de treino e teste, evitando problemas de *overfitting*. Essa

aproximação é mais visível com o aumento do número de épocas de treino. O comportamento semelhante das curvas mostra que o modelo consegue fazer previsões precisas tanto nos dados de treino quanto nos dados de teste.

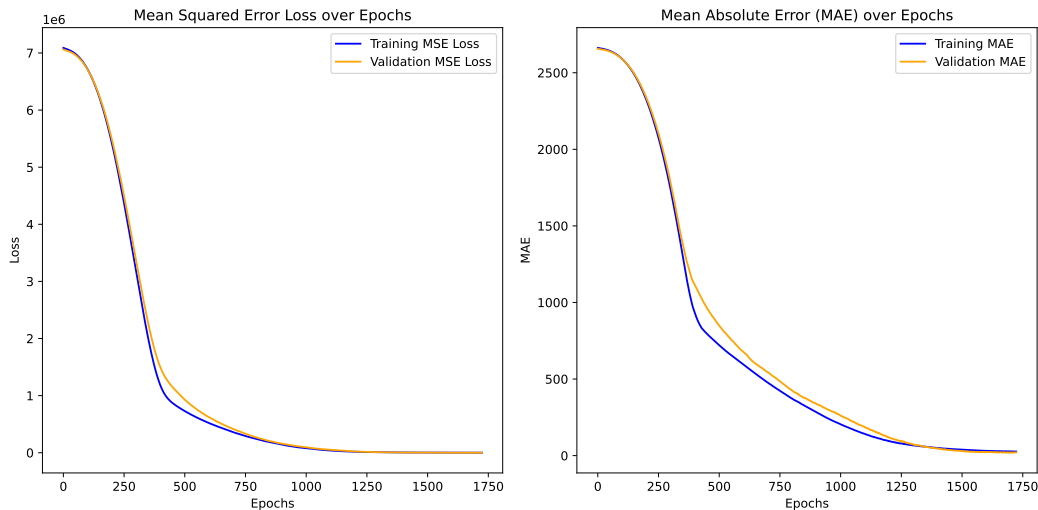


Figura 4.6: Gráfico do cálculo do MSE e do MAE por cada época

4.6 Erro de previsão

Para determinar a precisão do modelo, foi calculado o Coeficiente de Determinação (R^2). Esta métrica quantifica a precisão da previsão do modelo e varia de 0 a 1. O valor 0 significa que o modelo não representa nenhuma variabilidade nos dados e o valor 1 representa a total variabilidade nos dados, ajustando-se perfeitamente. Quanto mais próximo de 1, melhor o modelo se está a ajustar aos dados e o resultado obtido para o modelo desenvolvido neste projeto foi de 0,75 representando assim um bom ajuste.

Perante a análise das métricas de avaliação é possível observar que existem pequenos erros de previsão, mas nada de significativo. Estes erros podem ocorrer devido à natureza dos dados. Os dados utilizados, seguem padrões simples e lineares e o modelo consegue ajustar bem esses valores, o que resulta em pequenos erros de previsão. Outro fator que pode influenciar os resultados é o volume de dados utilizados, sendo que o *dataset* utilizado conta com poucos dados e isso pode ter dificultado o processo de aprendizagem do modelo.

Para interpretar o resultado do modelo foi criado o gráfico presente na Figura, 4.7. Este gráfico representa a comparação entre os valores previstos e os valores reais. Como é possível de analisar, os valores calculados pelo modelo não diferem muito da realidade, o que significa que o modelo está realizando previsões aproximadas

da realidade. Os valores para o Y_{pred} e Y_{test} , podem ser verificado na Tabela 4.4.

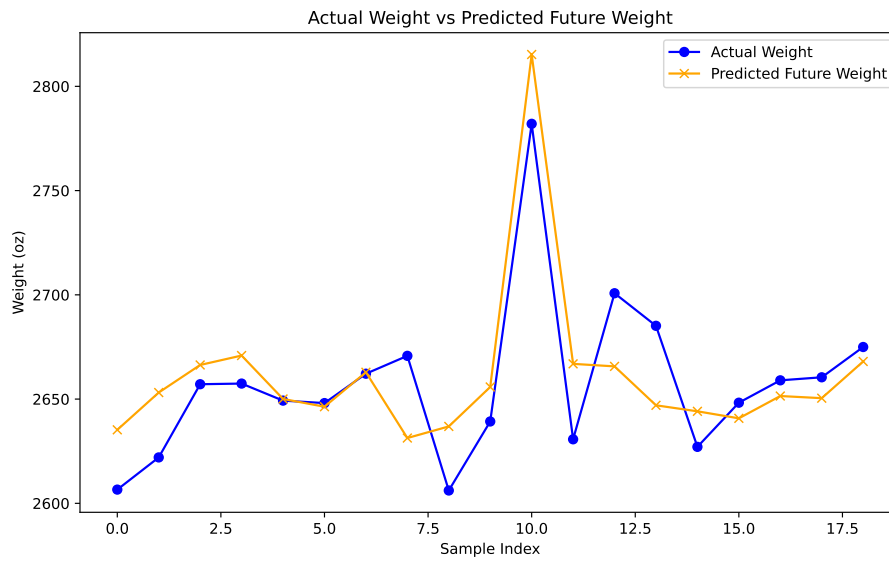


Figura 4.7: Gráfico da comparação entre o peso atual e o peso previsto

Y_test	Y_Pred
2606.58285714	2635.2566
2622.01714286	2653.1375
2657.12571429	2666.3628
2657.44	2670.8787
2649.26285714	2650.0603
2648.08	2646.3376
2662.16571429	2662.87
2670.73714286	2631.3225
2606.15428571	2636.8816
2639.26285714	2655.8125
2782.02857143	2815.2756
2630.71428571	2666.8938
2700.76571429	2665.6875
2685.14285714	2647.0164
2627.08571429	2644.1301
2648.26285714	2640.7712
2658.99428571	2651.466
2660.45142857	2650.4348
2674.97714286	2668.0398

Tabela 4.4: Valores das variáveis Y_pred e Y_test.

4.7 Discussão de resultados

O objetivo principal deste trabalho, é a criação dum modelo de previsão do peso de uma pessoa num curto prazo, a partir do seu comportamento nos dias anteriores ao da previsão. Esse comportamento inclui a ingestão de calorias, as calorias gastas em repouso (associadas à TMB) e as calorias gastas durante a atividade física.

Os resultados experimentais obtidos permitem responder às questões de investigação definidas na Secção 1.4.

(Q1): É viável desenvolver modelos preditivos capazes de estimar de forma precisa e imediata as consequências da ingestão de calorias e da prática de exercício físico no peso corporal no curto prazo?

Estes valores mostram que é viável desenvolver modelos preditivos capazes de estimar as consequências da ingestão de calorias e da prática de exercício físico no peso corporal. Os valores apresentados anteriormente mostra que o modelo prevê corretamente, tendo em conta o impacto que o estilo de vida tem no peso no curto espaço de tempo.

A utilização de um modelo de previsão permite a qualquer pessoa avaliar o seu comportamento imediato em termos de consumo de alimentos e exercício físico no seu peso nos dias seguintes. Pode também ser utilizado para antecipar o seu comportamento. Por exemplo, a pessoa pode evitar o consumo de um determinado alimento porque prevê o seu impacto negativo nos objetivos de controlo de peso no curto prazo. Da mesma forma, no que concerne ao exercício físico, a antecipação permite planejar a intensidade do exercício no próprio dia para que os objetivos de peso possam ser atingidos nos dias consequentes. Dentro da área da saúde e da nutrição estes modelos podem suportar a tomada de decisão dos profissionais.

(Q2): Que técnicas de pré-processamento de dados devem ser consideradas para melhorar a qualidade e eficácia dos modelos preditivos que procuram estimar as consequências imediatas da ingestão de calorias e da prática de exercício físico no peso corporal, incluindo fatores como normalização, tratamento de dados em falta, seleção de características relevantes?

As etapas de pré-processamento utilizadas, como a seleção de características relevantes, tratamento dos dados em falta e transformação de variáveis, mostraram-se relevantes para melhorar a qualidade e eficácia dos modelos preditivos. A escolha das variáveis de entrada permitiu selecionar os atributos que tinham uma maior importância para a previsão, o que contribuiu para a precisão e eficácia no desempenho do modelo. Assim como o tratamento de dados, que contribuiu para a remoção de dados irregulares. A seguir, foi possível gerar novos atributos essenciais para o desenvolvimento do modelo, como foi o caso do atributo 'TMB'. A função *Harris and Benedict* foi utilizada para calcular a TMB. Esta escolha permitiu incorporar aspetos fisiológicos individuais ao modelo, considerando as diferenças metabólicas que ocorrem num curto espaço de tempo. O deslocamento temporal, através da introdução de novas colunas, permitiu incluir o efeito do comportamento passado da pessoa, como, por exemplo, a ingestão de calorias em vários dias anteriores à data da previsão.

(Q3): Qual é o desempenho do modelo de previsão do peso corporal tendo em conta a ingestão de calorias e a prática de exercício físico?

A escolha da rede neuronal para a arquitetura do modelo e o ajuste dos hiperparâmetros foi essencial para a minimização do erro de previsão. A configuração da rede com um número reduzido de neurónios permitiu adequar o modelo ao tamanho reduzido do *dataset*, permitindo maximizar o desempenho do modelo utilizando um número reduzido de registos limitado.

A avaliação do modelo de previsão resultou num coeficiente de determinação (R²) de 0,75, revelando a capacidade de generalização do modelo. A capacidade de generalização também foi constatada pelos resultados experimentais para várias épocas de treino. Os resultados mostraram uma redução do MSE com o aumento do

número de épocas de treino, equivalente para o *dataset* de validação e treino. Como é possível de observar na Figura 4.6.

Capítulo 5

Conclusão

Ao longo deste trabalho, explorámos a viabilidade e a utilidade de desenvolver modelos preditivos capazes de estimar as consequências da ingestão de calorias e da prática de exercícios no peso corporal. Os resultados apresentados anteriormente indicam que tais modelos têm o potencial de prever com precisão o impacto do estilo de vida no peso a curto prazo.

Além da análise do impacto do estilo de vida no peso, é essencial considerar a importância do peso corporal na saúde geral de um indivíduo. O excesso de peso e a obesidade estão associados a uma série de problemas de saúde já discutidos. Portanto, compreender e monitorizar o peso corporal é crucial para a prevenção e o tratamento dessas condições.

A nutrição desempenha um papel fundamental na saúde e no bem-estar. A qualidade e a quantidade dos alimentos consumidos têm um impacto direto na composição corporal e na saúde geral. Portanto, o desenvolvimento de modelos ML que auxiliem na avaliação e no planeamento de padrões alimentares saudáveis é de importância.

Neste contexto, os modelos de ML surgem como ferramentas poderosas para a análise e previsão de padrões de comportamento relacionados à alimentação e ao exercício físico. A capacidade desses modelos de prever com precisão o impacto das escolhas de estilo de vida no peso corporal pode ser utilizada para auxiliar os indivíduos a tomar decisões mais informadas sobre a sua saúde e bem-estar.

Este trabalho não apenas demonstra o potencial dos modelos de ML na área da saúde e da nutrição, mas também destaca a importância contínua da pesquisa e do

desenvolvimento nesta área para promover um estilo de vida saudável e auxiliar na prevenção de doenças.

5.1 Limitações e desafios

As limitações deste trabalho estão relacionadas com o volume de dados disponíveis para alimentar o modelo. O processo de recolha de dados é crucial para a eficácia do modelo, uma vez que os utilizadores teriam de registar os seus comportamentos ao longo de um período prolongado, o que pode ser uma exigência bastante desafiadora.

Além disso, existe o risco de que os comportamentos dos utilizadores mudem ao longo do tempo, o que poderia levar à criação de um modelo obsoleto. Este problema é conhecido como *Concept Drift*, onde as relações entre as variáveis do modelo modificam-se, tornando as previsões do modelo menos precisas ou até mesmo inválidas [Zenisek et al., 2019]. Infelizmente, este cenário não foi contemplado na metodologia deste trabalho, o que pode comprometer a robustez e a utilidade do modelo proposto em ambientes dinâmicos.

Outra preocupação diz respeito à privacidade dos dados. Uma vez que lidamos com informações pessoais dos utilizadores, nem todos se sentirão confortáveis em partilhar esses dados. A recolha de dados pessoais pode levantar preocupações éticas e legais, o que pode afetar a aceitação e a adoção do sistema pelos utilizadores.

Além disso, a métrica de peso, sendo uma parte fundamental do modelo proposto, pode variar ao longo do dia. Portanto, é essencial que a recolha de dados seja sistematizada e realizada de maneira consistente para garantir a precisão das informações registadas.

Por fim, a obtenção de dados sobre calorias consumidas e gastas também apresenta desafios. Para uma quantificação precisa, pode ser necessário o uso de métodos de medição precisos, como balanças para determinar as quantidades de alimentos consumidos e smartwatches para monitorizar a atividade física. Estes dispositivos podem facilitar a recolha de dados, mas também aumentam a complexidade e os custos do sistema proposto.

5.2 Trabalho futuro

Com o objetivo de aprimorar a precisão do modelo, seria interessante incorporar variáveis contextuais que influenciem o peso, como o nível de atividade física, qualidade da dieta, hábitos de sono e fatores ambientais. A incorporação dessas variáveis pode levar a um entendimento mais abrangente e personalizado dos padrões de ganho ou perda de peso, tornando os modelos mais robustos e aplicáveis a uma gama mais ampla de cenários.

Pode também ser explorada a integração de dados biométricos, como a composição corporal, marcadores hormonais e a utilização de tecnologia mais avançada, como dispositivos que monitorizam dados fisiológicos em tempo real, para fornecer uma fonte rica de informação que possa ser analisada por profissionais de saúde.

Referências

- [Abhari et al., 2019] Abhari, S., Safdari, R., Azadbakht, L., Lankarani, K. B., Kahlori, S. R. N., Honarvar, B., Abhari, K., Ayyoubzadeh, S. M., Karbasi, Z., Zakerabasali, S., and Jalilpiran, Y. (2019). A Systematic Review of Nutrition Recommendation Systems: With Focus on Technical Aspects. *Journal of Biomedical Physics and Engineering*, 0(0):591–602. [Citado na página 1]
- [Agapito et al., 2016] Agapito, G., Calabrese, B., Guzzi, P. H., Cannataro, M., Simeoni, M., Caré, I., Lamprinoudi, T., Fuiano, G., and Pujia, A. (2016). Dietos: A recommender system for adaptive diet monitoring and personalized food suggestion. In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–8. IEEE. [Citado na página 25]
- [Ahmed et al., 2020] Ahmed, M., Seraj, R., and Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8). [Citado na página 10]
- [Albuquerque et al., 2017] Albuquerque, D., Nóbrega, C., Manco, L., and Padez, C. (2017). The contribution of genetics and environment to obesity. *British medical bulletin*, 123(1):159–173. [Citado na página 2]
- [Alimoradi et al., 2020] Alimoradi, Z., Golboni, F., Griffiths, M. D., Broström, A., Lin, C.-Y., and Pakpour, A. H. (2020). Weight-related stigma and psychological distress: A systematic review and meta-analysis. *Clinical Nutrition*, 39(7):2001–2013. [Citado nas páginas 18 e 19]
- [Alkhalaf et al., 2022] Alkhalaf, M., Yu, P., Shen, J., and Deng, C. (2022). A review of the application of machine learning in adult obesity studies. *Applied Computing and Intelligence*, 2(1):32–48. [Citado na página 3]
- [Almeida, 2020] Almeida, L. B. (2020). Multilayer perceptrons. In *Handbook of Neural Computation*, pages C1–2. CRC Press. [Citado na página 11]
- [Alves and de Oliveira Cunha, 2020] Alves, G. M. and de Oliveira Cunha, T. C. (2020). A importância da alimentação saudável para o desenvolvimento humano. *Humanas Sociais & Aplicadas*, 10(27):46–62. [Citado na página 28]

- [Anthanont and Jensen, 2016] Anthanont, P. and Jensen, M. D. (2016). Does basal metabolic rate predict weight gain? *The American journal of clinical nutrition*, 104(4):959–963. [Citado na página 21]
- [Badillo et al., 2020a] Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., and Zhang, J. D. (2020a). An introduction to machine learning. *Clinical pharmacology & therapeutics*, 107(4):871–885. [Citado na página 9]
- [Badillo et al., 2020b] Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., and Zhang, J. D. (2020b). An introduction to machine learning. *Clinical Pharmacology & Therapeutics*, 107(4):871–885. [Citado na página 10]
- [Baguley et al., 2022] Baguley, B. J., Dalla Via, J., Fraser, S. F., Daly, R. M., and Kiss, N. (2022). Effectiveness of combined nutrition and exercise interventions on body weight, lean mass, and fat mass in adults diagnosed with cancer: a systematic review and meta-analysis. *Nutrition Reviews*, 81(6):625–646. [Citado na página 19]
- [Barber and Franks, 2021] Barber, T. M. and Franks, S. (2021). Obesity and polycystic ovary syndrome. *Clinical endocrinology*, 95(4):531–541. [Citado na página 20]
- [Bashar, 2019] Bashar, D. A. (2019). Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence and Capsule Networks*, 1(2):73–82. [Citado na página 10]
- [Bellicha et al., 2021] Bellicha, A., van Baak, M. A., Battista, F., Beaulieu, K., Blundell, J. E., Busetto, L., Carraça, E. V., Dicker, D., Encantado, J., Ermolao, A., et al. (2021). Effect of exercise training on weight loss, body composition changes, and weight maintenance in adults with overweight or obesity: An overview of 12 systematic reviews and 149 studies. *Obesity Reviews*, 22:e13256. [Citado na página 28]
- [Bendavid et al., 2021] Bendavid, I., Lobo, D. N., Barazzoni, R., Cederholm, T., Coëffier, M., de van der Schueren, M., Fontaine, E., Hiesmayr, M., Laviano, A., Pichard, C., et al. (2021). The centenary of the harris–benedict equations: How to assess energy requirements best? recommendations from the espen expert group. *Clinical nutrition*, 40(3):690–701. [Citado na página 22]
- [Blüher, 2020] Blüher, M. (2020). Metabolically healthy obesity. *Endocrine reviews*, 41(3):bnaa004. [Citado na página 1]

- [Bouchard and Pérusse, 2020] Bouchard, C. and Pérusse, L. (2020). Genetics of obesity: family studies. In *The genetics of obesity*, pages 79–92. CRC press. [Citado na página 19]
- [Brannon et al., 2018] Brannon, L., Updegraff, J. A., and Feist, J. (2018). *Health psychology: An introduction to behavior and health*. Cengage Learning. [Citado nas páginas 3 e 20]
- [Brownlee, 2018] Brownlee, J. (2018). What is the difference between a batch and an epoch in a neural network. *Machine learning mastery*, 20. [Citado na página 15]
- [Buoncrisiano et al., 2021] Buoncrisiano, M., Williams, J., Simmonds, P., Nurk, E., Ahrens, W., Nardone, P., Rito, A. I., Rutter, H., Bergh, I. H., Starc, G., et al. (2021). Socioeconomic inequalities in overweight and obesity among 6-to 9-year-old children in 24 countries from the world health organization european region. *Obesity Reviews*, 22:e13213. [Citado na página 3]
- [Cao et al., 2018] Cao, W., Wang, X., Ming, Z., and Gao, J. (2018). A review on neural networks with random weights. *Neurocomputing*, 275:278–287. [Citado na página 12]
- [Carnauba et al., 2017] Carnauba, R. A., Baptistella, A. B., Paschoal, V., and Hübscher, G. H. (2017). Diet-induced low-grade metabolic acidosis and clinical outcomes: a review. *Nutrients*, 9(6):538. [Citado nas páginas 3 e 20]
- [Cecchini et al., 2010] Cecchini, M., Sassi, F., Lauer, J. A., Lee, Y. Y., Guajardo-Barron, V., and Chisholm, D. (2010). Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. *The Lancet*, 376(9754):1775–1784. [Citado na página 17]
- [Cena and Calder, 2020] Cena, H. and Calder, P. C. (2020). Defining a healthy diet: Evidence for the role of contemporary dietary patterns in health and disease. *Nutrients*, 12(2). [Citado nas páginas 19 e 28]
- [Chapman et al., 1999] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (1999). The crisp-dm user guide. In *4th CRISP-DM SIG Workshop in Brussels in March*, volume 1999. sn. [Citado na página 27]
- [Chapman et al., 2000] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9(13):1–73. [Citado nas páginas ix e 28]
- [Chebli et al., 2018] Chebli, A., Djebbar, A., and Marouani, H. F. (2018). Semi-supervised learning for medical application: A survey. In *2018 International Conference on Applied Smart Systems (ICASS)*, pages 1–9. [Citado na página 10]

- [Cheong and Freene, 2018] Cheong, L. and Freene, N. (2018). Exploring the patient journey in weight loss: A social network analysis. *Journal of Community Medicine and Health Education*, 8:1–1. [Citado na página 2]
- [Chi et al., 2021] Chi, X., Liang, K., Chen, S.-T., Huang, Q., Huang, L., Yu, Q., Jiao, C., Guo, T., Stubbs, B., Hossain, M. M., Yeung, A., Kong, Z., and Zou, L. (2021). Mental health problems among chinese adolescents during the covid-19: The importance of nutrition and physical activity. *International Journal of Clinical and Health Psychology*, 21(3):100218. [Citado na página 18]
- [Chicco et al., 2021] Chicco, D., Warrens, M. J., and Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, 7:e623. [Citado na página 16]
- [CHRISBOW, 2018] CHRISBOW (2018). 2018 calorie, exercise and weight changes. [Citado na página 33]
- [Cockerham et al., 2017] Cockerham, W. C., Hamby, B. W., and Oates, G. R. (2017). The social determinants of chronic disease. *American journal of preventive medicine*, 52(1):S5–S12. [Citado na página 2]
- [Conner and Norman, 2017] Conner, M. and Norman, P. (2017). Health behaviour: Current issues and challenges. [Citado na página 20]
- [Dastres and Soori, 2021] Dastres, R. and Soori, M. (2021). Artificial neural network systems. *International Journal of Imaging and Robotics (IJIR)*, 21(2):13–25. [Citado na página 10]
- [de Souza Pena et al., 2022] de Souza Pena, V., Gonçalves, A. C. R., Vieira, I. R., de Sousa, M. R., de Souza, A. C. D., La Croix, L. M. d. O., Fernandes, B. B., and da Cunha Gonçalves, S. J. (2022). Uma análise sobre as características da síndrome dos ovários policísticos: uma revisão de literatura. *Revista Eletrônica Acervo Médico*, 4:e9996–e9996. [Citado na página 20]
- [de Witt Huberts et al., 2013] de Witt Huberts, J. C., Evers, C., and de Ridder, D. T. (2013). Double trouble: restrained eaters do not eat less and feel worse. *Psychology & Health*, 28(6):686–700. [Citado na página 19]
- [Dhanachandra et al., 2015] Dhanachandra, N., Manglem, K., and Chanu, Y. J. (2015). Image segmentation using k -means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764–771. Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015,

- Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India. [Citado na página 10]
- [Di Renzo et al., 2020] Di Renzo, L., Gualtieri, P., Pivari, F., Soldati, L., Attinà, A., Cinelli, G., Leggeri, C., Caparello, G., Barrea, L., Scerbo, F., et al. (2020). Eating habits and lifestyle changes during covid-19 lockdown: an italian survey. *Journal of translational medicine*, 18(1):1–15. [Citado na página 19]
- [Error, 2016] Error, M. A. (2016). The mean absolute error. *Retrieved September, 19:2016*. [Citado na página 14]
- [Fairburn, 2008] Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press. [Citado na página 19]
- [Ferdowsy et al., 2021] Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I., and Habib, M. T. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences*, 2:100053. [Citado na página 23]
- [Forbush et al., 2020] Forbush, K. T., Song, Q., Tay, L., Gould, S. R., Chapa, D. A., Cushing, C. C., and Ptomey, L. T. (2020). Do differences between individuals who are healthy weight or overweight on self-report measures of disinhibited eating and restrained eating reflect reality or item “bias”? *Psychological Assessment*, 32(6):553. [Citado na página 18]
- [Frayn and Knäuper, 2018] Frayn, M. and Knäuper, B. (2018). Emotional eating and weight in adults: a review. *Current Psychology*, 37:924–933. [Citado na página 19]
- [Garrard, 2020] Garrard, J. (2020). Health sciences literature review made easy. [Citado na página 2]
- [geeksforgeeks, 2023] geeksforgeeks (2023). MI | underfitting and overfitting. [Citado nas páginas ix e 17]
- [Hamdan, 2018] Hamdan, M. K. (2018). *VHDL auto-generation tool for optimized hardware acceleration of convolutional neural networks on FPGA (VGT)*. PhD thesis, Iowa State University. [Citado na página 12]
- [Han and Weiss, 2021] Han, J. C. and Weiss, R. (2021). Obesity, metabolic syndrome and disorders of energy balance. In *Sperling pediatric endocrinology*, pages 939–1003. Elsevier. [Citado na página 20]
- [Han, 2021] Han, M. (2021). E-bayesian estimations of parameter and its evaluation standard: E-mse (expected mean square error) under different loss functions.

- Communications in Statistics-Simulation and Computation*, 50(7):1971–1988. [Citado na página 14]
- [Harrop et al., 2021] Harrop, E. N., Mensinger, J. L., Moore, M., and Lindhorst, T. (2021). Restrictive eating disorders in higher weight persons: A systematic review of atypical anorexia nervosa prevalence and consecutive admission literature. *International Journal of Eating Disorders*, 54(8):1328–1357. [Citado na página 2]
- [Haththotuwa et al., 2020] Haththotuwa, R. N., Wijeyaratne, C. N., and Senarath, U. (2020). Worldwide epidemic of obesity. In *Obesity and obstetrics*, pages 3–8. Elsevier. [Citado na página 17]
- [Hayden, 2022] Hayden, J. (2022). *Introduction to health behavior theory*. Jones & Bartlett Learning. [Citado na página 19]
- [Hill et al., 2020] Hill, J. O., Pagliassotti, M. J., and Peters, J. C. (2020). Nongenetic determinants of obesity and body fat topography. In *The genetics of obesity*, pages 35–48. CRC Press. [Citado na página 19]
- [Horta et al., 2023] Horta, B. L., Rollins, N., Dias, M. S., Garcez, V., and Pérez-Escamilla, R. (2023). Systematic review and meta-analysis of breastfeeding and later overweight or obesity expands on previous study for world health organization. *Acta Paediatrica*, 112(1):34–41. [Citado na página 3]
- [Hosseini and Padhy, 2019] Hosseini, S. A. and Padhy, R. K. (2019). Body image distortion. [Citado na página 20]
- [Iwendi et al., 2020] Iwendi, C., Khan, S., Anajemba, J. H., Bashir, A. K., and Noor, F. (2020). Realizing an efficient iomt-assisted patient diet recommendation system through machine learning model. *IEEE access*, 8:28462–28474. [Citado na página 23]
- [Jabbar and Khan, 2015] Jabbar, H. and Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 70(10.3850):978–981. [Citado na página 17]
- [Jagim et al., 2018] Jagim, A. R., Camic, C. L., Kisiolek, J., Luedke, J., Erickson, J., Jones, M. T., and Oliver, J. M. (2018). Accuracy of resting metabolic rate prediction equations in athletes. *The Journal of Strength & Conditioning Research*, 32(7):1875–1881. [Citado na página 22]
- [Johannink et al., 2019] Johannink, T., Bahl, S., Nair, A., Luo, J., Kumar, A., Loskyll, M., Ojea, J. A., Solowjow, E., and Levine, S. (2019). Residual reinforcement learning for robot control. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6023–6029. [Citado na página 10]

- [Juul et al., 2018] Juul, F., Martinez-Steele, E., Parekh, N., Monteiro, C. A., and Chang, V. W. (2018). Ultra-processed food consumption and excess weight among us adults. *British Journal of Nutrition*, 120(1):90–100. [Citado na página 19]
- [Kanal, 2003] Kanal, L. N. (2003). Perceptron. In *Encyclopedia of Computer Science*, pages 1383–1385. [Citado na página 11]
- [Kansal et al., 2018] Kansal, T., Bahuguna, S., Singh, V., and Choudhury, T. (2018). Customer segmentation using k-means clustering. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pages 135–139. [Citado na página 10]
- [Karaboga, 2010] Karaboga, D. (2010). Artificial bee colony algorithm. *scholarpedia*, 5(3):6915. [Citado na página 24]
- [Katte, 2018] Katte, T. (2018). Recurrent neural network and its various architecture types. *International Journal of Research and Scientific Innovation*, 5(3):124–129. [Citado na página 11]
- [Khanna et al., 2022] Khanna, D., Peltzer, C., Kahar, P., and Parmar, M. S. (2022). Body mass index (bmi): a screening tool analysis. *Cureus*, 14(2). [Citado na página 21]
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. [Citado na página 15]
- [Klaassen et al., 2017] Klaassen, G., Zelle, D. M., Navis, G. J., Dijkema, D., Bemelmann, F. J., Bakker, S. J., and Corpeleijn, E. (2017). Lifestyle intervention to improve quality of life and prevent weight gain after renal transplantation: Design of the active care after transplantation (act) randomized controlled trial. *Bmc nephrology*, 18:1–13. [Citado na página 19]
- [Kolotkin and Andersen, 2017] Kolotkin, R. L. and Andersen, J. R. (2017). A systematic review of reviews: exploring the relationship between obesity, weight loss and health-related quality of life. *Clinical obesity*, 7(5):273–289. [Citado na página 2]
- [Konttinen, 2020] Konttinen, H. (2020). Emotional eating and obesity in adults: the role of depression, sleep and genes. *Proceedings of the Nutrition Society*, 79(3):283–289. [Citado na página 19]
- [Lample and Chaplot, 2017] Lample, G. and Chaplot, D. S. (2017). Playing fps games with deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). [Citado na página 10]

- [Le et al., 2017] Le, L. K.-D., Barendregt, J. J., Hay, P., and Mihalopoulos, C. (2017). Prevention of eating disorders: A systematic review and meta-analysis. *Clinical psychology review*, 53:46–58. [Citado na página 2]
- [Leipold et al., 2018] Leipold, N., Madenach, M., Schäfer, H., Lurz, M., Terzimehic, N., Groh, G., Böhm, M., Gedrich, K., and Krcmar, H. (2018). Nutrilize a personalized nutrition recommender system: an enable study. *HealthRecSys@ RecSys*, 2216:24–29. [Citado na página 25]
- [Li et al., 2012] Li, J., Cheng, J.-h., Shi, J.-y., and Huang, F. (2012). Brief introduction of back propagation (bp) neural network algorithm and its improvement. In Jin, D. and Lin, S., editors, *Advances in Computer Science and Information Engineering*, pages 553–558, Berlin, Heidelberg. Springer Berlin Heidelberg. [Citado na página 23]
- [Li et al., 2021] Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*. [Citado na página 11]
- [Lighton, 2018] Lighton, J. R. (2018). *Measuring metabolic rates: a manual for scientists*. Oxford University Press. [Citado na página 21]
- [Lin et al., 2020] Lin, C.-Y., Imani, V., Broström, A., Huus, K., Björk, M., Hodges, E. A., and Pakpour, A. H. (2020). Psychological distress and quality of life in iranian adolescents with overweight/obesity: Mediating roles of weight bias internalization and insomnia. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 25(6):1583–1592. [Citado nas páginas 18 e 19]
- [Lin et al., 2017] Lin, W.-Q., Wang, H., Yuan, L.-X., Li, B., Jing, M.-J., Luo, J.-L., Tang, J., Ye, B.-K., and Wang, P.-X. (2017). The unhealthy lifestyle factors associated with an increased risk of poor nutrition among the elderly population in china. *The journal of nutrition, health & aging*, 21:943–953. [Citado na página 19]
- [Liu et al., 2018] Liu, G., Dhana, K., Furtado, J. D., Rood, J., Zong, G., Liang, L., Qi, L., Bray, G. A., DeJonge, L., Coull, B., et al. (2018). Perfluoroalkyl substances and changes in body weight and resting metabolic rate in response to weight-loss diets: a prospective study. *PLoS medicine*, 15(2):e1002502. [Citado na página 21]
- [Lopez-Jimenez et al., 2022] Lopez-Jimenez, F., Almahmeed, W., Bays, H., Cuevas, A., Di Angelantonio, E., le Roux, C. W., Sattar, N., Sun, M. C., Wittert, G., Pinto, F. J., et al. (2022). Obesity and cardiovascular disease: mechanistic insights and management strategies. a joint position paper by the world heart federation and world obesity federation. *European journal of preventive cardiology*, 29(17):2218–2237. [Citado na página 3]

- [MacLean et al., 2017] MacLean, P. S., Blundell, J. E., Mennella, J. A., and Battersham, R. L. (2017). Biological control of appetite: a daunting complexity. *Obesity*, 25:S8–S16. [Citado na página 19]
- [Marcason, 2015] Marcason, W. (2015). What are the components to the mind diet? *Journal of the Academy of Nutrition and Dietetics*, 115(10):1744. [Citado na página 19]
- [Marengo et al., 2018] Marengo, D., Longobardi, C., Fabris, M. A., and Settanni, M. (2018). Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. *Computers in Human Behavior*, 82:63–69. [Citado na página 2]
- [Marsland, 2011] Marsland, S. (2011). *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC. [Citado na página 10]
- [McKinley, 2017] McKinley, N. M. (2017). Ideal weight/ideal women: Society constructs the female. In *Weighty issues*, pages 97–115. Routledge. [Citado na página 18]
- [Molnar, 2020] Molnar, C. (2020). *Interpretable machine learning*. Lulu. com. [Citado na página 9]
- [Moradi et al., 2021] Moradi, M., Mozaffari, H., Askari, M., and Azadbakht, L. (2021). Association between overweight/obesity with depression, anxiety, low self-esteem, and body dissatisfaction in children and adolescents: a systematic review and meta-analysis of observational studies. *Critical Reviews in Food Science and Nutrition*, 62(2):555–570. [Citado na página 2]
- [Morales and Escalante, 2022] Morales, E. F. and Escalante, H. J. (2022). A brief introduction to supervised, unsupervised, and reinforcement learning. In *Biosignal processing and classification using computational learning and intelligence*, pages 111–129. Elsevier. [Citado na página 10]
- [Nagari and Inayati, 2020] Nagari, S. S. and Inayati, L. (2020). Implementation of clustering using k-means method to determine nutritional status. *J. Biometrika dan Kependud*, 9(1):62. [Citado na página 24]
- [Pearl and Puhl, 2018] Pearl, R. L. and Puhl, R. M. (2018). Weight bias internalization and health: a systematic review. *Obesity reviews*, 19(8):1141–1163. [Citado na página 20]
- [Pellegrini et al., 2020] Pellegrini, M., Cioffi, I., Evangelista, A., Ponzio, V., Goitre, I., Ciccone, G., Ghigo, E., and Bo, S. (2020). Effects of time-restricted feeding on body weight and metabolism. a systematic review and meta-analysis. *Reviews in endocrine and metabolic disorders*, 21:17–33. [Citado na página 19]

- [Peng et al., 2013] Peng, X., Zhou, C., Hepburn, D. M., Judd, M. D., and Siew, W. H. (2013). Application of k-means method to pattern recognition in on-line cable partial discharge monitoring. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(3):754–761. [Citado na página 10]
- [Phelan et al., 2015] Phelan, S. M., Burgess, D. J., Yeazel, M. W., Hellerstedt, W. L., Griffin, J. M., and van Ryn, M. (2015). Impact of weight bias and stigma on quality of care and outcomes for patients with obesity. *obesity reviews*, 16(4):319–326. [Citado na página 19]
- [Piaggi, 2019] Piaggi, P. (2019). Metabolic determinants of weight gain in humans. *Obesity*, 27(5):691–699. [Citado na página 22]
- [Powell et al., 2020] Powell, B. K. M., Machalek, D., and Quah, T. (2020). Real-time optimization using reinforcement learning. *Computers & Chemical Engineering*, 143:107077. [Citado na página 10]
- [Powell-Wiley et al., 2021] Powell-Wiley, T. M., Poirier, P., Burke, L. E., Després, J.-P., Gordon-Larsen, P., Lavie, C. J., Lear, S. A., Ndumele, C. E., Neeland, I. J., Sanders, P., et al. (2021). Obesity and cardiovascular disease: a scientific statement from the american heart association. *Circulation*, 143(21):e984–e1010. [Citado na página 3]
- [Purbasari et al., 2021] Purbasari, A., Rinawan, F. R., Zulianto, A., Susanti, A. I., and Komara, H. (2021). Crisp-dm for data quality improvement to support machine learning of stunting prediction in infants and toddlers. In *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pages 1–6. IEEE. [Citado na página 27]
- [Ramachandran et al., 2017] Ramachandran, P., Zoph, B., and Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*. [Citado na página 12]
- [Rasamoelina et al., 2020] Rasamoelina, A. D., Adjailia, F., and Sinčák, P. (2020). A review of activation function for artificial neural network. In *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMII)*, pages 281–286. IEEE. [Citado na página 13]
- [Rask-Andersen et al., 2019] Rask-Andersen, M., Karlsson, T., Ek, W. E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nature communications*, 10(1):339. [Citado na página 19]

- [Raut et al., 2018] Raut, M., Prabhu, K., Fatehpuria, R., Bangar, S., and Sahu, S. (2018). A personalized diet recommendation system using fuzzy ontology. *Int. J. Eng. Sci. Invention*, 7(3):51–55. [Citado na página 24]
- [Ravanelli et al., 2020] Ravanelli, M., Zhong, J., Pascual, S., Swietojanski, P., Monteiro, J., Trmal, J., and Bengio, Y. (2020). Multi-task self-supervised learning for robust speech recognition. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6989–6993. [Citado na página 10]
- [Reinders et al., 2017] Reinders, I., Visser, M., and Schaap, L. (2017). Body weight and body composition in old age and their relationship with frailty. *Current Opinion in Clinical Nutrition & Metabolic Care*, 20(1):11–15. [Citado na página 20]
- [Rosenblatt, 2021] Rosenblatt, F. (2021). The perceptron: A probabilistic model for information storage and organization (1958). [Citado na página 11]
- [Rostami et al., 2022] Rostami, M., Oussalah, M., and Farrahi, V. (2022). A novel time-aware food recommender-system based on deep learning and graph clustering. *IEEE Access*, 10:52508–52524. [Citado na página 24]
- [Roza and Shizgal, 1984] Roza, A. M. and Shizgal, H. M. (1984). The harris benedict equation reevaluated: resting energy requirements and the body cell mass. *The American journal of clinical nutrition*, 40(1):168–182. [Citado na página 22]
- [Rubino et al., 2022] Rubino, D. M., Greenway, F. L., Khalid, U., O’Neil, P. M., Rosenstock, J., Sørrig, R., Wadden, T. A., Wizert, A., Garvey, W. T., Arauz-Pacheco, C., et al. (2022). Effect of weekly subcutaneous semaglutide vs daily liraglutide on body weight in adults with overweight or obesity without diabetes: the step 8 randomized clinical trial. *Jama*, 327(2):138–150. [Citado na página 20]
- [Rucklidge and Kaplan, 2016] Rucklidge, J. J. and Kaplan, B. J. (2016). Nutrition and mental health. *Clinical Psychological Science*, 4(6):1082–1084. [Citado na página 18]
- [Safaei et al., 2021] Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., and Shapi’i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in biology and medicine*, 136:104754. [Citado na página 3]
- [Santanasto et al., 2017] Santanasto, A. J., Goodpaster, B. H., Kritchevsky, S. B., Miljkovic, I., Satterfield, S., Schwartz, A. V., Cummings, S. R., Boudreau, R. M., Harris, T. B., and Newman, A. B. (2017). Body composition remodeling and

- mortality: the health aging and body composition study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 72(4):513–519. [Citado na página 18]
- [Sathya et al., 2013] Sathya, R., Abraham, A., et al. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2):34–38. [Citado na página 10]
- [Schmidhuber, 2015] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117. [Citado na página 16]
- [Schröer et al., 2021] Schröer, C., Kruse, F., and Gómez, J. M. (2021). A systematic literature review on applying crisp-dm process model. *Procedia Computer Science*, 181:526–534. CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020. [Citado na página 27]
- [Silva, 2022] Silva, D. d. O. (2022). Otimização de hiper-parâmetros de algoritmos de machine learning aplicados no contexto de análise de risco de crédito. [Citado na página 12]
- [Silventoinen and Konttinen, 2020] Silventoinen, K. and Konttinen, H. (2020). Obesity and eating behavior from the perspective of twin and genetic research. *Neuroscience & biobehavioral reviews*, 109:150–165. [Citado na página 19]
- [Soares et al., 2020] Soares, G. V. D., Soares, C. V. D., de Medeiros, T. K. F., and dos Santos, E. B. (2020). Distúrbios fisiológicos relacionados à glândula tireoide: uma revisão literária. *Research, Society and Development*, 9(7):e376974258–e376974258. [Citado na página 20]
- [Sonawane and Patil, 2014] Sonawane, J. S. and Patil, D. R. (2014). Prediction of heart disease using multilayer perceptron neural network. In *International conference on information communication and embedded systems (ICICES2014)*, pages 1–6. IEEE. [Citado na página 22]
- [Sun et al., 2022] Sun, M., Jia, W., Chen, G., Hou, M., Chen, J., and Mao, Z.-H. (2022). Improved wearable devices for dietary assessment using a new camera system. *Sensors*, 22(20):8006. [Citado na página 17]
- [Sung et al., 2019] Sung, H., Siegel, R. L., Torre, L. A., Pearson-Stuttard, J., Islami, F., Fedewa, S. A., Goding Sauer, A., Shuval, K., Gapstur, S. M., Jacobs, E. J.,

- et al. (2019). Global patterns in excess body weight and the associated cancer burden. *CA: a cancer journal for clinicians*, 69(2):88–112. [Citado na página 18]
- [Sutton and Barto, 2018] Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press. [Citado na página 10]
- [Svendsen et al., 2020] Svendsen, M. T., Bak, C. K., Sørensen, K., Pelikan, J., Riddersholm, S. J., Skals, R. K., Mortensen, R. N., Maindal, H. T., Bøggild, H., Nielsen, G., et al. (2020). Associations of health literacy with socioeconomic position, health risk behavior, and health status: a large national population-based survey among danish adults. *BMC public health*, 20(1):1–12. [Citado na página 1]
- [Tsoumakas, 2019] Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, 52(1):441–447. [Citado na página 10]
- [Turgeon and Rioux, 2011] Turgeon, S. L. and Rioux, L.-E. (2011). Food matrix impact on macronutrients nutritional properties. *Food Hydrocolloids*, 25(8):1915–1924. 25 years of Advances in Food Hydrocolloid Research. [Citado na página 19]
- [Van Strien, 2018] Van Strien, T. (2018). Causes of emotional eating and matched treatment of obesity. *Current diabetes reports*, 18:1–8. [Citado na página 19]
- [Verduci et al., 2022] Verduci, E., Di Profio, E., Fiore, G., and Zuccotti, G. (2022). Integrated approaches to combatting childhood obesity. *Annals of Nutrition and Metabolism*, 78(Suppl. 2):8–19. [Citado na página 20]
- [Wang et al., 2020] Wang, Q., Ma, Y., Zhao, K., and Tian, Y. (2020). A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, pages 1–26. [Citado na página 14]
- [Weber and Achananuparp, 2016] Weber, I. and Achananuparp, P. (2016). Insights from machine-learned diet success prediction. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pages 540–551. World Scientific. [Citado na página 25]
- [Westmoreland et al., 2016] Westmoreland, P., Krantz, M. J., and Mehler, P. S. (2016). Medical complications of anorexia nervosa and bulimia. *The American journal of medicine*, 129(1):30–37. [Citado na página 18]
- [Wu and Feng, 2018] Wu, Y.-c. and Feng, J.-w. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656. [Citado na página 10]
- [Xuan et al., 2020] Xuan, W., Williams, K., and Peat, J. K. (2020). *Health science research: A handbook of quantitative methods*. Routledge. [Citado na página 2]

- [Yang and Shami, 2020] Yang, L. and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316. [Citado na página 15]
- [Zenisek et al., 2019] Zenisek, J., Holzinger, F., and Affenzeller, M. (2019). Machine learning based concept drift detection for predictive maintenance. *Computers & Industrial Engineering*, 137:106031. [Citado na página 46]
- [Zhang et al., 2023] Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2023). *Dive into deep learning*. Cambridge University Press. [Citado nas páginas ix, 13 e 14]
- [Zhang et al., 2019] Zhang, H., Zhang, L., and Jiang, Y. (2019). Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. [Citado na página 16]
- [Zhao et al., 2024] Zhao, F.-Y., Liu, J.-E., Fang, X.-M., Chen, L., Liang, J.-G., Liu, Y., Li, R.-L., and Chen, H.-R. (2024). Effects of a 12-week exercise-based intervention on weight management in overweight or obese breast cancer survivors: a randomized controlled trial. *Supportive Care in Cancer*, 32(2):98. [Citado na página 19]
- [Zhao et al., 2021] Zhao, H., Zhang, X., Xu, Y., Gao, L., Ma, Z., Sun, Y., and Wang, W. (2021). Predicting the risk of hypertension based on several easy-to-collect risk factors: A machine learning method. *Frontiers in Public Health*, 9. [Citado nas páginas 4 e 23]
- [Zheng and Ruggiero, 2017] Zheng, Z. and Ruggiero, K. (2017). Using machine learning to predict obesity in high school students. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2132–2138. [Citado na página 3]
- [Zhou, 2021] Zhou, Z.-H. (2021). *Machine learning*. Springer nature. [Citado na página 10]