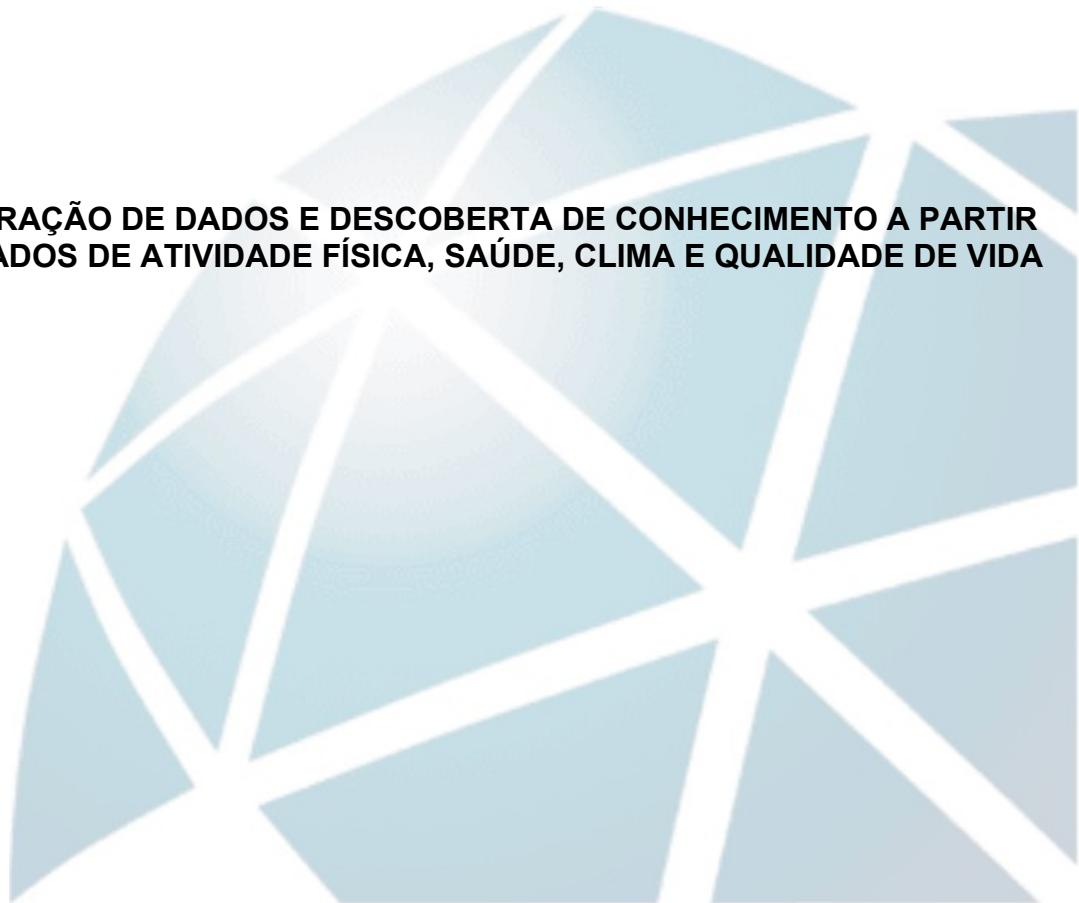

**PÓS-GRADUAÇÃO EM DESENVOLVIMENTO HUMANO E TECNOLOGIAS
(ÁREA DE TECNOLOGIAS NAS DINÂMICAS CORPORAIS)**

Rafael Macedo Sulino

**MINERAÇÃO DE DADOS E DESCOBERTA DE CONHECIMENTO A PARTIR
DE DADOS DE ATIVIDADE FÍSICA, SAÚDE, CLIMA E QUALIDADE DE VIDA**



Rio Claro-SP
2020



Rafael Macedo Sulino

MINERAÇÃO DE DADOS E DESCOBERTA DE CONHECIMENTO A PARTIR DE DADOS DE ATIVIDADE FÍSICA, SAÚDE, CLIMA E QUALIDADE DE VIDA

Tese de doutorado apresentada ao Instituto de Biociências da Universidade Estadual Paulista, campus de Rio Claro, como requisito parcial para obtenção do título de Doutor em Desenvolvimento Humano e Tecnologias, área de concentração Tecnologias nas Dinâmicas Corporais.

Orientador:

Prof. Dr. Carlos Norberto Fischer

Coorientador:

Prof. Dr. Wellington Roberto Gomes Carvalho

Rio Claro-SP
2020

S949m	<p>Sulino, Rafael Macedo Mineração de dados e descoberta de conhecimento a partir de dados de atividade física, saúde, clima e qualidade de vida / Rafael Macedo Sulino. -- Rio Claro, 2020 158 p. : il., tabs., fotos</p> <p>Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Rio Claro Orientador: Carlos Norberto Fischer Coorientador: Wellington Roberto Gomes de Carvalho</p> <p>1. Educação Física. 2. Mineração de dados. 3. Exercícios físicos. 4. Tecnologia. I. Título.</p>
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA TESE: Mineração de dados e descoberta de conhecimento a partir de dados de atividade física, saúde, clima e qualidade de vida.

AUTOR: RAFAEL MACEDO SULINO

ORIENTADOR: CARLOS NORBERTO FISCHER

COORIENTADOR: WELLINGTON ROBERTO GOMES DE CARVALHO

Aprovado como parte das exigências para obtenção do Título de Doutor em DESENVOLVIMENTO HUMANO E TECNOLOGIAS, área: Tecnologias nas Dinâmicas Corporais pela Comissão Examinadora:

Prof. Dr. CARLOS NORBERTO FISCHER (Participação Virtual)
Departamento de Estatística, Matemática Aplicada e Computação / UNESP - Instituto de Geociências e Ciências Exatas Rio Claro- SP

Prof. Dr. AFONSO ANTONIO MACHADO (Participação Virtual)
Departamento de Educação Física / UNESP - Instituto de Biociências de Rio Claro / SP

Prof. Dr. ALEXANDRE GABARRA DE OLIVEIRA (Participação Virtual)
Departamento de Educação Física / UNESP - Instituto de Biociências de Rio Claro - SP

Prof. Dr. ANDRE FRANCESCHI DE ANGELIS (Participação Virtual)
Faculdade de Tecnologia / UNICAMP

Prof. Dr. GUSTAVO LIMA ISLER (Participação Virtual)
Faculdades Integradas Claretianas / Unidade Rio Claro / SP

Rio Claro, 28 de outubro de 2020

Dedico este trabalho aos meus pais, que sempre me incentivaram e nunca mediram esforços para me proporcionar tudo o que foi necessário para alcançar meus objetivos.

AGRADECIMENTOS

Concluir uma tese de doutorado não é tarefa fácil, exigindo do pesquisador muita dedicação e horas de estudo, além de feriados e fins de semana dedicados ao trabalho. Ao longo desta trajetória, conta-se, inevitavelmente, com o apoio e incentivo de muitas pessoas. Manifesto aqui os meus sinceros agradecimentos a todos aqueles que, de alguma forma, contribuíram para a concretização desta tese.

Ao meu orientador, Prof. Dr. Carlos Norberto Fischer, pela atenção, disponibilidade, ensinamentos, e pela pertinência de suas críticas e sugestões.

Ao meu coorientador, Prof. Dr. Wellington Roberto Gomes de Carvalho, pelas contribuições, amizade e incentivo, desde os tempos da graduação no Instituto Federal em Muzambinho.

Aos membros da banca, Prof. Dr. Afonso Antônio Machado, Prof. Dr. Alexandre Gabarra de Oliveira, Prof. Dr. André Franceschi de Angelis e Prof. Dr. Gustavo Lima Isler, pelas contribuições e ensinamentos.

Aos professores e colaboradores do campus de Rio Claro da Universidade Estadual Paulista, por proporcionarem um ambiente acadêmico do mais alto nível.

Aos meus pais, meus primeiros mestres, responsáveis acima de tudo pela formação do meu caráter.

À minha companheira e parceira em todos os momentos, Caroline Botan Ribeiro, pela paciência, carinho e compreensão.

Aos familiares, amigos e todos aqueles que de alguma forma, direta ou indiretamente, contribuíram para a concretização desta importante etapa da minha vida.

“Nenhuma decisão sensata pode ser tomada sem que se leve em conta o mundo não apenas como ele é, mas como ele virá a ser (ASIMOV, 1975).”

RESUMO

A produção de dados, dos mais diversos tipos, é um fenômeno cada vez mais presente em nosso dia a dia. Neste contexto, a prática de atividade física vem cada vez mais se beneficiando com tal fenômeno, por meio da disponibilidade cada vez maior de dispositivos e aplicativos móveis capazes de registrar e armazenar os dados da atividade, tais como coordenadas de GPS, frequência cardíaca, altimetria e velocidade, como o popular *Strava*. Além disso, o acesso a dados públicos, como os climáticos, é cada vez mais farto e facilitado, e tais dados podem ser combinados com registros de atividade física. No entanto, muitas vezes, os dados produzidos não recebem um tratamento e análise que possam trazer novos conhecimentos, de modo a contribuir de alguma forma para melhorar a prática da atividade física, o que proporcionaria melhor desempenho, maior conforto ou mesmo prevenção de lesões. A escassez de tratamento e análise de tais dados justifica o desenvolvimento deste trabalho, que teve por objetivo propor uma plataforma computacional capaz de auxiliar na descoberta de conhecimentos relevantes, por meio da análise de dados de ciclismo e corrida de rua, registrados especificamente pelo aplicativo *Strava*, associados a dados climáticos, tais como temperatura, umidade relativa do ar, velocidade e direção do vento, que possam ser utilizados tanto por praticantes de atividade física quanto por profissionais que orientam a prática de atividade física, de forma a auxiliar na tomada de decisões para o planejamento de uma prática mais eficiente e segura. Por meio de um estudo experimental contando com 1059 participantes, sendo 844 do sexo masculino e 215 do sexo feminino, que praticaram 168.409 atividades de ciclismo e 72.009 atividades de corrida de rua, registradas por meio do aplicativo *Strava*, foi possível aplicar o processo de Mineração de Dados, usando a técnica chamada Árvore de Decisão, possibilitando identificar diversas relações entre variáveis de desempenho, clima e qualidade de vida. A técnica empregada nas análises dos tipos de dados tratados produziu resultados interessantes, viabilizando, dessa forma, a construção do protótipo de uma plataforma, denominada *Fitness Tools*, por meio da qual os profissionais de Educação Física serão capazes de realizar o cadastro de seus clientes e o registro de dados, bem como análises referentes às atividades desenvolvidas. A partir dos dados registrados, o profissional passa a contar com ferramentas de análise por meio da técnica Árvore de Decisão, capazes de fornecer informações que possam auxiliá-lo no planejamento, acompanhamento e ajustes relacionados ao treinamento esportivo. Este estudo mostrou o potencial de utilização da Mineração de Dados na descoberta de conhecimentos que possam ser utilizados para a tomada de decisões quanto à prática de atividade física de forma mais segura e eficiente. A Mineração de Dados oferece uma opção relativamente flexível, adaptável a características específicas (região, clima, tipo de terreno), podendo ser utilizada para um único atleta, por meio de ajustes levando em conta seu próprio histórico (como volume, intensidade, características do terreno e clima), ou para grandes grupos de atletas, como os populares grupos de corrida.

Palavras-chave: Atividade física. Plataforma computacional para coleta e análise de dados. Mineração de Dados de atividade física. Extração de conhecimento de dados de ciclismo e corrida de rua. *Strava*.

ABSTRACT

The production of data, of the most diverse types, is a phenomenon increasingly present in our day to day. In this context, the practice of physical activity has increasingly benefited from this phenomenon, through the rise of the availability of devices and mobile applications capable of recording and storing activity data, such as GPS coordinates, heart rate, altimetry, and speed, as the popular Strava. In addition, access to public data, such as climate data, is abundant and facilitated, and such data can be combined with physical activity records. However, often the data produced do not receive a treatment and analysis that can bring new knowledge, in order to contribute in some way to improve the practice of physical activity, which would provide better performance, greater comfort, or even prevention of injuries. The lack of treatment and analysis of such data justifies the development of this work, which aimed to propose a computational platform capable of assisting in the discovery of relevant knowledge, through the analysis of cycling and running data, recorded specifically by the Strava, associated with weather data such as temperature, relative air humidity, wind speed, and direction, that can be used both by physical activity practitioners and professionals who guide the practice of it, in order to assist in decision-making for the planning of a more efficient and safer practice. Through an experimental study with 1059 participants, 844 men and 215 women, who practiced 168,409 cycling and 72,009 running activities, registered through the Strava, it was possible to apply the Data Mining process, using the technique called Decision Tree, allowing to identify several relationships between performance, climate, and quality of life variables. The technique used in the analysis of the types of processed data produced interesting results, thus enabling the construction of a prototype platform, called Fitness Tools, through which Physical Education professionals will be able to register their clients and record data, as well as analyses regarding the activities developed. From the registered data, the professional starts to rely on analysis tools using the Decision Tree technique, able to provide information that can assist him in planning, monitoring, and adjustments related to sports training. This study showed the potential of using Data Mining to discover knowledge that can be used to make decisions about the practice of physical activity in a safer and more efficient way. Data Mining offers a relatively flexible option, adaptable to specific characteristics (region, climate, type of terrain), and can be used for a single athlete, through adjustments taking into account his own history (such as volume, intensity, terrain characteristics, and weather), or for large groups of athletes, such as the popular racing groups.

Keywords: Physical activity. Computational platform for data collection and analysis. Mining physical activity data. Knowledge extraction from cycling and running data. Strava.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de árvore de decisão.....	27
Figura 2 – Etapas do processo de descoberta de conhecimento.....	29
Figura 3 – Integração entre aparelho de academia e <i>Apple Watch</i>	40
Figura 4 – Sistema <i>Zwift</i>	40
Figura 5 – Sistema <i>Polar V800</i>	42
Figura 6 – <i>Apple Watch</i> com ECG.....	43
Figura 7 – Aplicativos mais populares na categoria “Saúde e Atividade Física” em Fevereiro de 2020.....	44
Figura 8 – Aplicativo <i>Strava</i>	44
Figura 9 – Distribuição da amostra em 136 países.....	48
Figura 10 – Opção de revogação de acesso aos dados do <i>Strava</i>	49
Figura 11 – Página de autenticação do <i>Strava</i>	51
Figura 12 – Configurações do servidor VPS.....	52
Figura 13 – Fluxograma de coleta de dados.....	53
Figura 14 – Comparação entre a direção do vento e a direção do percurso.....	61
Figura 15 – Tendência relacionada ao desempenho.....	62
Figura 16 – Distribuição dos tipos de atividades, em percentual.....	64
Figura 17 – Exemplo de seleção das variáveis mais relevantes.....	66
Figura 18 – Análise exploratória (ciclismo).....	70
Figura 19 – Histogramas (ciclismo).....	72
Figura 20 – Matriz de correlações entre as variáveis (ciclismo).....	74
Figura 21 – Árvore de decisão: influência do IMC sobre o desempenho.....	75
Figura 22 – Árvore de decisão: influência da velocidade do vento sobre a velocidade máxima.....	76
Figura 23 – Árvore de decisão: influência das direções do vento e do percurso sobre o desempenho.....	77
Figura 24 – Árvore de decisão: influência das direções do vento e do percurso em relação à velocidade máxima.....	78
Figura 25 – Árvore de decisão: influência da umidade relativa do ar sobre o <i>suffer score</i>	78
Figura 26 – Árvore de decisão: influência do perfil sobre o histórico de lesões.....	79

Figura 27 – Árvore de decisão: influência das zonas de intensidade sobre o histórico de lesões.....	80
Figura 28 – Árvore de decisão: influência do ganho de elevação sobre a percepção de desconforto.....	82
Figura 29 – Árvore de decisão: influência da temperatura sobre a percepção de desconforto.....	83
Figura 30 – Árvore de decisão: influência do perfil e do tempo decorrido sobre a percepção de desconforto.....	85
Figura 31 – Árvore de decisão: influência do histórico de lesões sobre a percepção de desconforto.....	86
Figura 32 – Árvore de decisão: influência do <i>score</i> do domínio físico de qualidade de vida sobre a intensidade.....	87
Figura 33 – Árvore de decisão: influência do <i>score</i> do domínio físico de qualidade de vida sobre a percepção de desconforto.....	88
Figura 34 – Análise exploratória (corrida de rua).....	90
Figura 35 – Histogramas (corrida de rua).....	91
Figura 36 – Matriz de correlações entre as variáveis (corrida de rua).....	92
Figura 37 – Árvore de decisão: influência do IMC sobre o desempenho.....	93
Figura 38 – Árvore de decisão: influência da temperatura sobre o desempenho.....	94
Figura 39 – Árvore de decisão: influência da umidade relativa do ar sobre o <i>suffer score</i>	95
Figura 40 – Árvore de decisão: influência da temperatura sobre o <i>suffer score</i>	96
Figura 41 – Árvore de decisão: influência da temperatura e umidade relativa do ar sobre o desempenho.....	97
Figura 42 – Árvore de decisão: influência da distância percorrida sobre o histórico de lesões.....	98
Figura 43 – Árvore de decisão: influência do ganho de elevação sobre o histórico de lesões.....	99
Figura 44 – Árvore de decisão: influência do <i>score</i> do domínio físico de qualidade de vida sobre o histórico de lesões.....	100
Figura 45 – Árvore de decisão: influência do perfil sobre o histórico de lesões.....	101
Figura 46 – Árvore de decisão: influência das zonas de intensidade sobre o histórico de lesões.....	102

Figura 47 – Árvore de decisão: influência das direções do vento e do percurso sobre o <i>suffer score</i>	103
Figura 48 – Autenticação de usuários da plataforma <i>Fitness Tools</i>	106
Figura 49 – Tela inicial – opção “Atletas”.....	107
Figura 50 – Exemplo de análise individual (mapa e altimetria).....	109
Figura 51 – Exemplo de análise individual (velocidade e frequência cardíaca).....	110
Figura 52 – Tela inicial – opção “Mineração de Dados”.....	111
Figura 53 – Ferramenta de Mineração de Dados.....	112
Figura 54 – Exemplo 1: influência do ganho de elevação na percepção de desconforto (ciclismo).....	113
Figura 55 – Exemplo 2: influência da distância percorrida sobre o histórico de lesões (corrida de rua).....	114

LISTA DE QUADROS

Quadro 1 – Domínios e facetas do instrumento WHOQOL Abreviado.....	36
Quadro 2 – Valores críticos associados ao grau de confiança na amostra.....	47
Quadro 3 – Descrição das variáveis obtidas por meio da API do <i>Strava</i>	55
Quadro 4 – Descrição das variáveis obtidas por meio da API do <i>World Weather Online</i>	56
Quadro 5 – Descrição das variáveis obtidas por meio da API do <i>LocationIQ</i>	56
Quadro 6 – Descrição das variáveis complementares.....	57
Quadro 7 – Valores de referência para as zonas de treinamento.....	60
Quadro 8 – Classificação do IMC.....	63
Quadro 9 – Classificação das zonas de conforto e desconforto térmico de acordo com a temperatura efetiva em função do vento (TEv) calculada.....	63

LISTA DE TABELAS

Tabela 1 – Caracterização da amostra.....	46
Tabela 2 – Caracterização das atividades realizadas (ciclismo).....	69
Tabela 3 – Caracterização das atividades realizadas (corrida de rua).....	89

LISTA DE EQUAÇÕES

(1) Cálculo do tamanho da amostra.....	46
(2) Índice de Massa Corporal.....	58
(3) Temperatura efetiva em função do vento.....	58
(4) WHOQOL abreviado: score domínio físico.....	59
(5) WHOQOL abreviado: score domínio psicológico.....	59
(6) WHOQOL abreviado: score domínio social.....	59
(7) WHOQOL abreviado: score domínio ambiental.....	59
(8) Frequência cardíaca máxima estimada.....	59
(9) Pace estimado.....	59
(10) Zona de intensidade Z1.....	60
(11) Zona de intensidade Z2.....	60
(12) Zona de intensidade Z3.....	60
(13) Zona de intensidade Z4.....	60
(14) <i>Bearing</i>	60

LISTA DE SIGLAS E ABREVIAÇÕES

ACSM	<i>American College of Sports Medicine</i> (Colégio Americano de Medicina do Esporte)
API	<i>Application Programming Interface</i> (interface de programação de aplicativos)
AVE	Acidente Vascular Encefálico
CEP	Comitê de Ética em Pesquisa
GPS	<i>Global Positioning System</i> (sistema de posicionamento global)
ICP-Brasil	Infraestrutura de Chaves Públicas Brasileiras
IMC	Índice de Massa Corporal
IP	<i>Internet Protocol</i> (endereço de identificação de redes utilizado em conexões na Internet)
KDD	<i>Knowledge Discover in Databases</i> (descoberta de conhecimento em bancos de dados)
MET	<i>Metabolic Equivalents</i> (equivalente metabólico)
NFC	<i>Near-field Communication</i> (protocolo de comunicação entre dispositivos eletrônicos)
OMS	Organização Mundial da Saúde
RAM	<i>Random Access Memory</i> (memória de acesso randômico)
RFE	<i>Redundant Feature Elimination</i> (eliminação de variáveis redundantes)
SSL	<i>Secure Socket Layer</i> (camada de conexão segura)
TCLE	Termo de Consentimento Livre e Esclarecido
VPS	<i>Virtual Private Server</i> (servidor virtual privado)

SUMÁRIO

1 INTRODUÇÃO.....	18
2 OBJETIVOS.....	22
3 REVISÃO DA LITERATURA.....	23
3.1 Bancos de dados e extração de conhecimento.....	23
3.1.1 Dado, informação e conhecimento.....	23
3.1.2 Sistemas de banco de dados.....	24
3.1.3 Mineração de Dados e descoberta de conhecimento.....	26
3.2 Atividade física, saúde e qualidade de vida.....	30
3.2.1 Atividade física.....	30
3.2.2 Planejamento da atividade física.....	32
3.2.3 Saúde e qualidade de vida.....	35
3.2.4 Influência do clima na prática de atividade física.....	37
3.3 Tecnologia e atividade física.....	39
4 MATERIAL E MÉTODOS.....	46
4.1 Amostra.....	46
4.2 Coleta e armazenamento de dados.....	50
4.3 Variáveis coletadas.....	54
4.4 Variáveis complementares.....	58
4.5 Tabelas de referência.....	63
4.6 Limpeza e preparação dos dados.....	63
4.7 Redução de variáveis.....	65
4.8 Mineração de Dados.....	66
5 ESTUDO EXPERIMENTAL.....	68
5.1 Ciclismo.....	68
5.1.1 Amostra.....	68
5.1.2 Análise exploratória e estatísticas.....	69
5.1.3 Mineração de Dados aplicada ao ciclismo.....	74
5.2 Corrida de rua.....	89
5.2.1 Amostra.....	89
5.2.2 Análise exploratória e estatísticas.....	90

5.2.3 Mineração de Dados aplicada à corrida de rua.....	93
5.3 Considerações finais sobre o estudo experimental.....	103
6 PLATAFORMA <i>FITNESS TOOLS</i>.....	105
6.1 Construção da plataforma.....	105
6.2 Segurança e privacidade.....	106
6.3 Entrada de dados.....	106
6.4 Análise de dados.....	108
7 CONSIDERAÇÕES FINAIS.....	115
REFERÊNCIAS.....	117
APÊNDICE A – TCLE (PORTUGUÊS).....	125
APÊNDICE B – TCLE (INGLÊS).....	128
APÊNDICE C – TCLE (ESPAÑOL).....	131
APÊNDICE D – CONVITE REDE SOCIAL DO STRAVA (PORTUGÊS).....	134
APÊNDICE E – CONVITE REDE SOCIAL DO STRAVA (INGLÊS).....	135
APÊNDICE F – CONVITE REDE SOCIAL DO STRAVA (ESPAÑOL).....	136
APÊNDICE G – CONVITE REDE SOCIAL <i>INSTAGRAM</i> (PORTUGÊS).....	137
APÊNDICE H – CONVITE REDE SOCIAL <i>INSTAGRAM</i> (INGLÊS).....	138
APÊNDICE I – ESPECIFICAÇÕES TÉCNICAS DA PLATAFORMA <i>FITNESS TOOLS</i>.....	139
APÊNDICE J – MODELO ENTIDADE RELACIONAMENTO DA PLATAFORMA <i>FITNESS TOOLS</i>.....	147
APÊNDICE H – QUESTIONÁRIO DE ANAMNESE.....	148
ANEXO A – PARECER CONSUSTANCIADO DO CEP.....	151
ANEXO B – DECISÃO CEP/IB 01/2019.....	154
ANEXO C – QUESTIONÁRIO WHOQOL ABREVIADO.....	156

1 INTRODUÇÃO

Do ponto de vista tecnológico, a década de 2010 trouxe consigo, com grande destaque, a popularização da “Internet das Coisas”, em que diversas formas de interação entre pessoas e dispositivos digitais resultam em um grande volume de dados (MADAKAM; RAMASWAMY; TRIPATHI, 2015; SETHI; SARANGI, 2017). Uma tendência recente é o desenvolvimento de sensores “vestíveis” (*wearables*), *smartphones* e uma série de outros equipamentos capazes de coletar dados e se comunicarem entre si, e estima-se que a quantidade de produtos desse tipo em uso já supera a quantidade de pessoas existentes no planeta, o que possibilita uma coleta expressiva de dados relevantes para a aplicação de algoritmos da chamada Mineração de Dados (AGGARWAL, 2015). No contexto da busca de uma vida saudável, tais dispositivos ganham notável importância ao possibilitarem a produção e compartilhamento de dados em tempo real. Dessa forma, temos a disposição diversos aplicativos que oferecem dietas nutricionais e programas de treinamento baseados em dados do indivíduo, como gênero, idade e massa corporal, além de seus objetivos, como o emagrecimento (OIKAWA, 2013). A evolução das tecnologias digitais no campo da saúde possibilita a exploração de forma mais profunda dos dados, de modo a rastrear informações que, no passado, poderiam ser obtidas somente por meio de exames médicos (FANTONI, 2016).

É notável o crescente aumento na quantidade de pessoas que utilizam algum aplicativo ou dispositivo para monitorar a prática de atividade física e saúde de modo geral. Estes recursos permitem monitorar o gasto calórico, histórico alimentar, frequência cardíaca em repouso e durante a prática de atividade física, tempo e qualidade do sono, entre muitas outras variáveis, abrangendo dessa forma os principais componentes para a manutenção e promoção de uma boa saúde, ou seja, boa alimentação, prática regular de atividade física e qualidade do sono.

No entanto, geralmente, tais dados não recebem tratamento e análise adequados, o que possibilitaria, tanto ao praticante quanto ao profissional da área de saúde, uma melhor tomada de decisões no sentido de planejar as atividades físicas de forma mais eficiente de acordo com o objetivo (emagrecimento, ganho de

performance, controle de diabetes, entre outros). Serrano et al. (2016) apontam que mais da metade dos aplicativos instalados em *smartphones* envolvem algo relacionado a controle de massa corporal, dieta ou exercício. Além disso, os dados coletados por meio destes aplicativos relacionados à saúde e atividade física, como abrangem milhões de usuários, ao serem estudados de forma metódica, oferecem potencial para importantes descobertas de conhecimento.

O processo de descoberta de conhecimento em bancos de dados (KDD) conta com muitos algoritmos de Mineração de Dados como, por exemplo, as chamadas Árvores de Decisão e os algoritmos de classificação. A Mineração de Dados é uma das etapas do KDD na qual, após o tratamento adequado dos dados, utilizam-se algoritmos específicos, conforme os tipos de dado e resultado desejados. Este processo de descoberta de conhecimento é capaz de revelar padrões desconhecidos, tendências de comportamento de determinada variável e muitas outras informações úteis. Com uma capacidade computacional cada vez mais elevada, a Mineração de Dados vêm sendo empregada em larga escala em diversas áreas, como mercado financeiro, medicina e mobilidade urbana.

Mesmo com a popularização de aplicativos e dispositivos capazes de monitorar a atividade física e parâmetros de saúde, produzindo uma farta quantidade de dados, muitos profissionais da área de Educação Física continuam utilizando métodos mais tradicionais, como formulários impressos e planilhas eletrônicas. Quando o profissional consegue ter acesso a recursos computacionais mais completos (por exemplo, por meio da aquisição de uma licença de *software*), ele encontra aplicativos voltados para a avaliação física de seus clientes (por exemplo, o *Physical Test*¹ e o *Trainer Metrics*²), de forma a automatizar alguns cálculos, facilitando a aplicação dos protocolos mais conhecidos, além de aplicativos que facilitam a prescrição do treinamento, por meio de modelos prontos que podem ser ajustados de acordo com as características de seus clientes (por exemplo, o *MFIT Personal*³ e o *FitSW*⁴). Tais aplicativos oferecem, ainda, ferramentas que auxiliam,

1 Disponível em <https://www.terrazul.com.br>

2 Disponível em <https://www.trainermetrics.com>

3 Disponível em <https://mfitpersonal.com.br>

4 Disponível em <https://www.fitsw.com>

por exemplo, no controle de frequência dos clientes às atividades propostas e gerenciamento financeiro.

Buscando na Internet especificamente ferramentas que possam facilitar a análise de dados obtidos por meio de dispositivos ou aplicativos móveis, como o *Strava*⁵, o que encontram-se são *websites* com dicas e tutoriais mostrando alguma análise específica, utilizando um pequeno conjunto de dados (FARCZADI, 2020; NASH, 2020). Alguns destes tutoriais compartilham exemplos de códigos de programação de computadores, geralmente incompreensíveis para profissionais que não sejam de áreas correlatas à Tecnologia da Informação. O acesso a recursos tecnológicos que possam facilitar a análise de dados, tornando-as intuitivas, de modo a auxiliar o profissional na tomada de decisões quanto ao planejamento do treinamento pode potencializar os resultados esperados.

Diante do exposto, justificam-se os esforços no sentido de se buscar novos conhecimentos a partir dos dados coletados por aplicativos e dispositivos de monitoramento de atividade física, por meio da Mineração de Dados. Dessa forma, o presente estudo, aprovado pelo Comitê de Ética em Pesquisa com seres humanos do Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP, campus de Rio Claro/SP, apresenta uma estrutura computacional que possibilitou a coleta e armazenamento de dados de ciclismo e corrida de rua, além de análises por meio de técnicas da Mineração de Dados. Tal estrutura viabilizou a elaboração de um estudo experimental, por meio do qual foi possível encontrar relações importantes entre diversas variáveis estudadas. Estes resultados mostram que a aplicação de tais técnicas pode ser bastante interessante em processos de análise dos tipos de dados coletados, de forma a viabilizar a construção de uma plataforma computacional que ofereça, tanto ao usuário quanto ao profissional, *feedback* e sugestões de interesse que possam auxiliar na tomada de decisões para o planejamento da prática de atividade física, levando-se em consideração fatores como o desempenho, saúde, qualidade de vida e influência de fatores climáticos.

5 Disponível em <https://www.strava.com>

A plataforma resultante do presente estudo, denominada *Fitness Tools*, é o protótipo de uma aplicação multiplataforma, disponível via Internet, por meio da qual o profissional de Educação Física será capaz de realizar o cadastro de seus clientes e também a integração de dados com o aplicativo *Strava*. A partir desta integração, o profissional passa a contar com ferramentas de análise por meio de algoritmos utilizados na Mineração de Dados, capazes de fornecer informações que possam auxiliá-lo no planejamento, acompanhamento e ajustes relacionados ao treinamento físico.

2 OBJETIVOS

O objetivo geral deste estudo consistiu em propor uma plataforma computacional capaz de promover a descoberta de conhecimentos relevantes, por meio da análise de dados coletados por aplicativos e dispositivos de monitoramento de atividade física além de dados climáticos obtidos por meio de requisição em interfaces de programação (APIs) de acesso público, que possa ser utilizada tanto por praticantes de atividade física quanto por profissionais que orientam tal prática, de forma a auxiliar na tomada de decisões para o planejamento de um treinamento mais eficiente e seguro.

Diante do contexto exposto, foram estabelecidos os seguintes objetivos específicos:

- a) Desenvolver uma estrutura capaz de receber o cadastro dos usuários participantes e autorização para acesso a seus dados de atividade física por meio da API do *Strava*⁶;
- b) Complementar os dados de atividade física com dados climáticos, de percepção de saúde e percepção da qualidade de vida;
- c) Aplicar técnicas da Mineração de Dados, objetivando a identificação de padrões que possam levar à descoberta de conhecimento relevante;
- d) Desenvolver uma plataforma computacional que seja flexível, possibilitando adaptações para considerar novos tipos de dados, inclusão de novas funcionalidades e novos métodos voltados para análise de dados e com tempo de vida relativamente longo.

⁶ Disponível em <https://developers.strava.com>

3 REVISÃO DA LITERATURA

Neste capítulo é apresentado um suporte teórico ao presente estudo, por meio de uma revisão da literatura existente. No primeiro tópico, são abordados os conceitos de *dado, informação e conhecimento*, bem como os processos computacionais empregados na extração de conhecimento em bancos de dados. No segundo, são revistos conceitos de atividade física, saúde e qualidade de vida, abordando também a influência de fatores climáticos. Por fim, no terceiro e último tópico, é apresentada a aplicação de tecnologias computacionais na área da saúde, especialmente na Educação Física.

3.1 Bancos de dados e extração de conhecimento

Interagimos a todo momento com diversos dispositivos e sistemas, resultando na produção dos mais variados tipos de dados (publicações em redes sociais, compartilhamento de fotos e vídeos, informações de geolocalização (GPS), transações financeiras, reserva em hotéis, locação de veículos, compras online, pesquisas em mecanismos de buscas, entre muitos outros). No entanto, tais dados por si só, em formato bruto, não são capazes de oferecer benefícios diretos, ou seja, devem ser tratados e processados de modo a possibilitar uma análise e extração de conhecimentos úteis.

3.1.1 Dado, informação e conhecimento

Nas primeiras civilizações, o ser humano já utilizava recursos para representar a realidade a sua volta. Ao observar as pinturas rupestres, é possível inferir sobre o modo de vida daqueles povos, a cultura, o ambiente em que viviam, ou seja, o homem pré-histórico já produzia dados, informações e conhecimentos.

Embora não exista um consenso quanto às definições de dado, informação e conhecimento, vários autores (DAVENPORT; KLAHR, 2012; DUTTA, 1997; MARSHALL, 1997) procuram apresentar as diferenças entre tais conceitos. Segundo Tuomi (1999), geralmente tais conceitos são tratados de forma hierárquica, na qual

dados são simples fatos que se tornam informação, se forem combinados com uma estrutura compreensível, enquanto a informação se torna conhecimento, se for inserida em um contexto. Dessa forma, uma informação é convertida em conhecimento quando podemos conectá-la a outras informações, de modo a avaliar e entender seu significado dentro de um contexto específico (SILVA, 2005). Neste sentido, os dados são pré-requisitos para a informação, que por sua vez é pré-requisito para o conhecimento.

Na área da tecnologia da informação, Lima e Alvares (2012) apresentam definições para dado, informação e conhecimento. Dado é definido como a menor partícula da informação que, de forma isolada, não possibilita uma decodificação clara para o entendimento humano, pois está fora de um contexto que favoreça sua compreensão. Informação é um conjunto de dados que possibilita a extração de algum significado, podendo ser utilizado para a obtenção de conhecimento. Conhecimento, por sua vez, está relacionado aos aspectos cognitivos que ocorrem na mente humana, envolvendo processos de captação, assimilação, associação, construção, desconstrução e reconstrução de conceitos.

Segundo Lima e Alvares (2012), o conhecimento comporta tanto um processo individual e mental quanto um conjunto de saberes que se desenvolvem e evoluem continuamente, concentrados em uma determinada área, ciência ou domínio da atividade humana.

Diante do enorme volume de dados que nos cerca, para que possamos extrair informações e produzir conhecimentos, contamos com os sistemas de gerenciamento de bancos de dados, tanto em versões pagas quanto gratuitas, os quais proporcionam grande capacidade de armazenamento e oferecem ferramentas que viabilizam a recuperação de conjuntos de informações, consequentemente possibilitando a produção do conhecimento.

3.1.2 Sistemas de banco de dados

Os sistemas de gerenciamento de bancos de dados se tornaram parte de nosso cotidiano. A todo momento nos deparamos com situações que, de alguma

maneira, envolvem a interação com algum destes sistemas. Por exemplo, quando realizamos uma transação financeira, como o pagamento de um boleto, seja por meio da agência bancária, caixa eletrônico ou aplicativo no computador ou *smartphone*, estamos interagindo com sistemas de banco de dados, assim como quando efetuamos uma compra online, ou mesmo em uma loja física (que possui um *software* de controle de estoque dos produtos, cadastro de clientes, etc.), quando contratamos um seguro, quando realizamos uma reserva em hotel, entre diversas outras situações.

Segundo Date (2004), um sistema de banco de dados é basicamente um sistema computadorizado para a manutenção de registros, ou seja, o equivalente eletrônico de um armário de arquivos. A finalidade de um sistema de banco de dados, de modo geral, é armazenar dados e permitir que o usuário possa buscar e atualizar tais dados, sempre que for necessário.

Além dos sistemas de gerenciamento de bancos de dados tradicionais (aqueles utilizados para armazenar registros alfanuméricos, como nomes, endereços, números de telefone, preços, quantidades em estoque), os avanços tecnológicos viabilizaram aplicações inovadoras, como os bancos de dados multimídia, capazes de armazenar imagens, vídeos e sons, os sistemas de informações geográficas, capazes de armazenar mapas e imagens de satélite, e os bancos de dados ativos em tempo real, utilizados em processos industriais, como linhas de produção, entre outros (ELMASRI; NAVATHE, 2005).

Os sistemas de banco de dados estão disponíveis nos mais variados dispositivos, desde pequenos computadores de mão, como *smartphones*, passando por computadores pessoais, servidores de empresas, até os maiores *mainframes*⁷. Dessa forma, o que vai determinar o tipo de sistema a ser utilizado (monousuário ou multiusuário; integrado ou compartilhado) é basicamente seu tamanho e finalidade (DATE, 2004).

⁷ *Mainframes* são computadores de grande porte, geralmente destinados a processar um volume muito grande de informações ou realizar operações de alta complexidade (LAUDON; LAUDON, 2014).

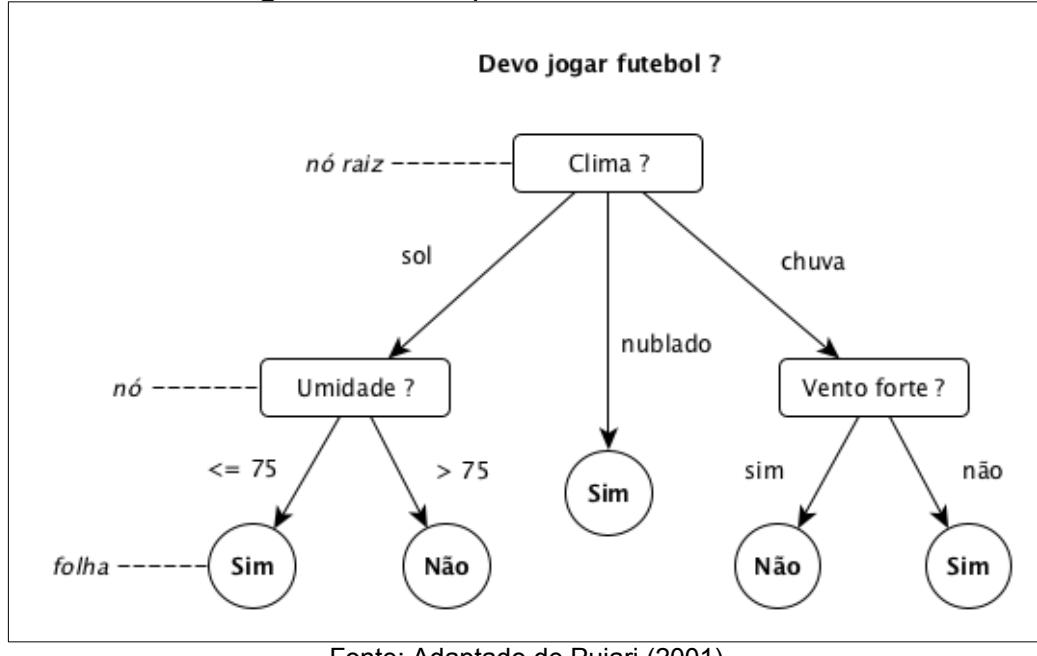
Com o avanço da capacidade computacional, com computadores pessoais se tornando capazes de realizar tarefas que antigamente eram possíveis somente em *mainframes*, somado a uma crescente produção de dados pelas mais diversas fontes e dos mais variados tipos, torna-se cada vez mais viável a exploração de bancos de dados para a descoberta de conhecimento.

3.1.3 Mineração de Dados e descoberta de conhecimento

Convivemos atualmente com uma imensurável quantidade de dados gerados a todo momento, por meio das mais variadas fontes. Muitas vezes esse grande volume de dados produzidos se torna inútil, devido à demanda sobre-humana necessária para analisar e extrair algum conhecimento. Na década de 1990, surgiram os primeiros processos informatizados capazes de realizar uma análise robusta de grandes volumes de dados, exclusivamente em ambiente acadêmico. Nas décadas seguintes, a utilização de sistemas para análises de bancos de dados saiu do ambiente acadêmico para ganhar espaço em áreas como mercado financeiro, controle de produção, medicina, entre outros (HAN; KAMBER; PEI, 2012), surgindo uma nova área de atuação profissional conhecida como Descoberta de Conhecimento em Bancos de Dados (KDD).

Dentre as etapas do KDD, destaca-se a Mineração de Dados, a qual emprega métodos voltados para identificar padrões nos dados analisados. Han, Kamber e Pei (2012) salientam a utilização de árvores de decisão como uma das técnicas mais eficientes da Mineração de Dados. As árvores de decisão são estruturas de classificação no formato da estrutura de dados denominada Árvore (Figura 1), na qual cada “nó” interno representa uma condição e cada “ramo” representa o caminho quanto ao teste de cada condição, finalizando com as “folhas”, que representam o resultado das decisões (CASTRO; FERRARI, 2016).

Figura 1 – Exemplo de árvore de decisão



Fonte: Adaptado de Pujari (2001)

Segundo Ye (2003), alguns fatores tornam as árvores de decisão especialmente atrativas: possuem uma representação intuitiva, o que torna os resultados mais fáceis de serem visualizados; são não paramétricas, adequadas para exploração e descoberta de conhecimento; os algoritmos voltados para sua criação são relativamente rápidos quando comparados aos de outros métodos e a acurácia é comparável a outros métodos de classificação.

Além das árvores de decisão, existem várias outras técnicas voltadas para a Mineração de Dados, as quais podem ser utilizadas por meio da implementação dos respectivos algoritmos nas mais diversas linguagens de programação de computadores, com destaque atualmente para *Python*⁸ e *R*⁹. É possível também utilizar ferramentas que facilitam a aplicação da Mineração de Dados por meio de interface gráfica, como o pacote *Weka*¹⁰, que é de acesso livre e implementa grande parte destes algoritmos.

Segundo Amo (2004), o processo de descoberta de conhecimento em bancos de dados (Figura 2) passa por uma sequência iterativa de etapas, iniciando-se com

8 Disponível em <https://www.python.org>

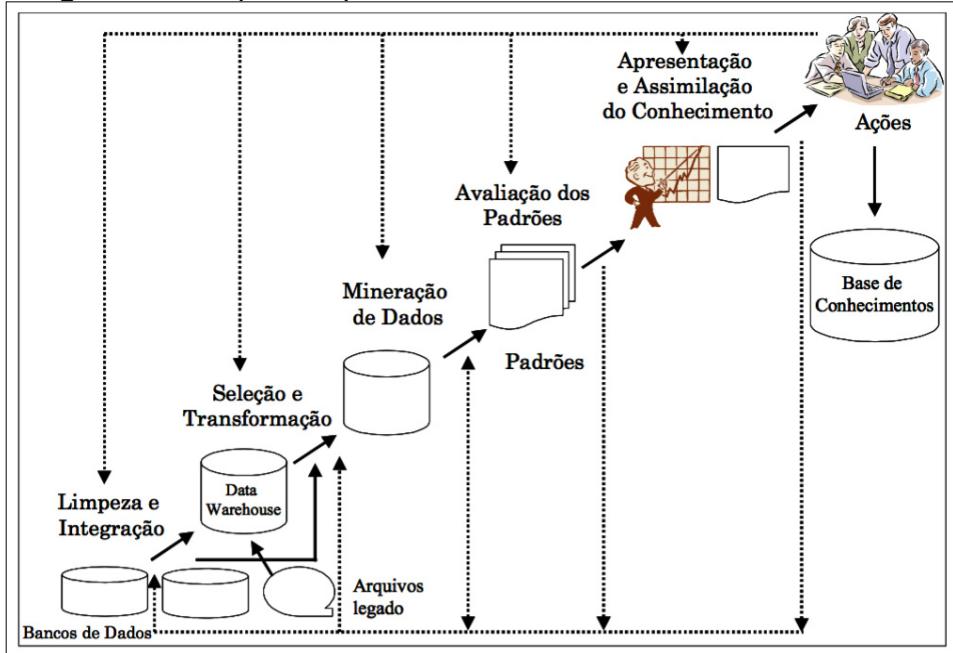
9 Disponível em <https://www.r-project.org>

10 Disponível em <https://www.cs.waikato.ac.nz/ml/weka/>

a preparação dos dados coletados (limpeza, integração, seleção e transformação), passando pela mineração desses dados transformados e finalizando com a avaliação e apresentação dos padrões encontrados (o conhecimento obtido). A fase de preparação se inicia com a limpeza, que consiste na eliminação de “ruídos” e registros inconsistentes. Em seguida, a integração possibilita a combinação de dados de diferentes fontes, quando pertinente, produzindo um único repositório. A preparação segue com a seleção dos atributos de interesse da pesquisa e a transformação de variáveis, na qual ocorrem os ajustes necessários para um formato apropriado, por exemplo, para a aplicação de algoritmos de Mineração de Dados. Ye (2003) destaca a importância desta etapa de preparação, pois, segundo o autor, esta consome entre 60% e 90% do tempo necessário para todo o processo e contribui de 75 a 90% para o sucesso do projeto. O processo de descoberta de conhecimento segue com a etapa de mineração dos dados, que consiste na aplicação de algoritmos para extrair padrões de possível interesse. Em seguida, ocorre a avaliação desses padrões, uma etapa de pós-processamento na qual busca-se a identificação de elementos relevantes, de acordo com os critérios estabelecidos pelo pesquisador. Por fim, a apresentação e assimilação do conhecimento tem por objetivo empregar técnicas de visualização e representação de informações, de forma que o resultado de todo o processo possa ser mais claramente compreendido (HAN; KAMBER; PEI, 2012).

Quando lidamos com um grande número de variáveis, nos deparamos com a necessidade de identificar aquelas que são realmente relevantes, bem como eliminar as redundantes. Segundo Ye (2003), a seleção de variáveis é um processo utilizado para encontrar um subconjunto delas a partir de um conjunto original, com o objetivo de reduzir a dimensão do conjunto de dados e proporcionar maior acurácia e velocidade aos algoritmos da Mineração de Dados.

Figura 2 – Etapas do processo de descoberta de conhecimento



Fonte: Côrtes, Porcaro e Lifschitz (2002)

A Mineração de Dados, considerada o ponto chave no processo de descoberta de conhecimento em bancos de dados, pode ser definida como um processo de proposição de várias consultas e extração de informações úteis, padrões e tendências, frequentemente desconhecidos, a partir de grande quantidade de dados armazenada (THURAISINGHAM, 1999). Segundo Côrtes, Porcaro e Lifschitz (2002), o conceito de Mineração de Dados se populariza cada vez mais como uma ferramenta de descoberta de informações capazes de revelar estruturas de conhecimento e guiar decisões em condições de certeza limitada. Trata-se de um processo cooperativo entre seres humanos e máquinas, visando explorar bancos de dados com o objetivo de extrair conhecimentos por meio de reconhecimento de padrões e relacionamento entre variáveis, utilizando-se técnicas comprovadamente confiáveis e validadas.

Banaee, Ahmed e Loutfi (2013) apontam três tipos de técnicas predominantes na Mineração de Dados aplicada, mais particularmente, a dados de saúde: predição, detecção de anomalias e tomada de decisões para diagnóstico. A predição é amplamente utilizada e consiste em auxiliar na identificação de eventos que ainda não ocorreram. Esta abordagem tem conquistado interesse na área de saúde por

auxiliar na decisão quanto a prognósticos. A detecção de anomalias consiste em identificar padrões incomuns, que não condizem com o comportamento esperado em um certo conjunto de dados. Detectar padrões incomuns em dados de saúde possibilita uma tomada de decisão mais precisa por parte do profissional de saúde.

Além das técnicas predominantes, podem ser utilizadas diversas outras, destacando-se os métodos baseados em regras e a análise de frequências. Os métodos baseados em regras consistem no reconhecimento de padrões com base em regras previamente estabelecidas. Já a análise de frequências pode ser empregada na avaliação de dados temporais, como a velocidade, variação de altitude e a frequência cardíaca durante a atividade física, possibilitando extrair padrões e tendências (BANAE; AHMED; LOUTFI, 2013).

3.2 Atividade física, saúde e qualidade de vida

De modo geral, a atividade física é frequentemente associada a uma boa saúde e, consequentemente, a uma boa qualidade de vida. No entanto, a prática de atividade física sem uma orientação e acompanhamento por profissionais, e sem levar em consideração alguns princípios básicos, como frequência, volume e intensidade, pode levar a inúmeros prejuízos tanto para a saúde quanto para a qualidade física, como lesões, além da não obtenção do objetivo almejado. Para que seja efetivamente benéfica, a prática de atividade física deve respeitar limites individuais, ter uma orientação profissional, além da importância de se entender e respeitar variáveis que possam afetar o sucesso da prática de atividade física, como os fatores climáticos.

3.2.1 Atividade física

A atividade física pode ser definida como qualquer movimento corporal produzido por músculos esqueléticos, que resulte em dispêndio de energia, geralmente mensurado em quilocalorias, enquanto o exercício físico é definido como um subconjunto da atividade física, composto por atividades realizadas de forma planejada, estruturada, repetitiva e com o objetivo de manter ou aprimorar a aptidão

física (BARBANTI, 2003; CASPERSEN; POWELL; CHRISTENSON, 1985). No dia a dia, podemos classificar as atividades físicas como ocupacionais, esportivas, cuidados domésticos, entre outras atividades. Como o presente estudo contém dados de atividade física de diferentes perfis de participantes, não necessariamente de natureza planejada, estruturada e repetitiva, optou-se por utilizar o termo “atividade física” para aquelas referentes à prática de atividades esportivas.

A prática regular de atividade física está associada a inúmeros benefícios (DAHN; AND, 2005; OJA et al., 2011; WARBURTON; BREDIN, 2017; WARBURTON; NICOL; BREDIN, 2006), tanto para a saúde física quanto mental, para homens e mulheres, sendo capaz de reduzir os riscos de acidente vascular encefálico (AVE), normalizar a pressão arterial, melhorar a sensibilidade a insulina, prevenir o desenvolvimento de doenças como diabetes, entre outros benefícios, além de desempenhar um papel importante no controle da massa corporal (GARBER et al., 2011).

Fang et al. (2019) estudaram os efeitos de um programa de atividade física com duração de 12 semanas no controle da massa corporal, aptidão física, estresse ocupacional, satisfação com o trabalho e qualidade de vida em trabalhadores sedentários. Os resultados do estudo apontam melhora significativa em todos os aspectos analisados. Arslan et al. (2019) estudaram os efeitos da prática de atividade física na qualidade do sono, satisfação com o trabalho e qualidade de vida em trabalhadores de escritórios, e concluíram que a prática regular de atividade física melhora tanto a satisfação com o trabalho quanto a qualidade de vida.

Segundo Bennell e Crossley (1996), a prática de atividade física de forma exaustiva, sem orientação profissional, ou de forma inadequada, pode contribuir significativamente para a ocorrência de lesões. A prática de atividade física deve ser realizada com acompanhamento profissional, respeitando-se algumas etapas fundamentais. Inicialmente, é necessária uma avaliação física, de forma que o profissional possa conhecer o estado de saúde e condicionamento físico do praticante, além de possíveis limitações, para que possa prescrever de forma adequada os parâmetros para a prática da atividade física desejada. Após a

prescrição do treinamento, é necessário um acompanhamento para os ajustes necessários e, ao final de um determinado período, realizar uma reavaliação para conhecer os resultados obtidos e planejar o próximo ciclo.

A prescrição do treinamento determinará, basicamente, parâmetros como o volume e intensidade da atividade física durante determinado período, de forma a garantir a eficiência do treinamento e a alcançar os objetivos do praticante, seja qual for (emagrecimento, desempenho esportivo, entre outros) (HEYWARD, 2013). Ou seja, o profissional vai determinar a periodicidade da prática, ritmo, velocidade, distância, entre outras variáveis, de acordo com princípios científicos do treinamento esportivo.

Após a prescrição do treinamento, durante o período da prática da atividade física, diversos fatores poderão e deverão ser observados, de forma a produzir informações que possibilitem adequar ou ajustar o treinamento, tais como o clima, a resposta fisiológica ao treinamento prescrito, quantidade e qualidade de sono do praticante, entre muitos outros fatores.

3.2.2 Planejamento da atividade física

Segundo Heyward (2013), alguns princípios básicos de treinamento se aplicam a qualquer tipo de programa de exercícios, seja qual for o objetivo almejado. (DANTAS, 2014) apresenta os seis princípios científicos do treinamento esportivo:

1) O princípio da individualidade biológica preconiza que cada indivíduo é único. A associação entre o genótipo (carga genética) e o fenótipo (tudo que é acrescido ao indivíduo a partir do nascimento) é responsável pela caracterização do indivíduo, de modo que gêmeos univitelinos (mesmo genótipo) terão experiências diferentes durante a vida, se tornando indivíduos diferentes (DANTAS, 2014).

2) O princípio da adaptação se refere à capacidade do corpo humano se adaptar a situações adversas, como variação de temperatura, situações que provocam emoções, variação de pressão, esforço físico, entre outras. Sempre que

ocorre um desequilíbrio da homeostase¹¹, o organismo procura reestabelecer o equilíbrio por meio de mecanismos compensatórios. As causas de desequilíbrio podem ser físicas, bioquímicas e mentais. O estresse físico é provocado pelo aumento da atividade física, de acordo com a intensidade do esforço. O estresse bioquímico ocorre devido à introdução de qualquer substância química no organismo, como álcool e fumo. Já o estresse mental pode ser provocado por ansiedade, angústia, entre outros fatores (DANTAS, 2014).

3) O princípio da sobrecarga diz respeito à necessidade de se aplicar cargas de intensidade crescentes para que ocorra progresso no treinamento. Deve-se respeitar um tempo de recuperação proporcional à intensidade e volume da atividade realizada, pois cargas maiores exigem maior tempo de recuperação. Normalmente, a sobrecarga é aplicada inicialmente ao volume e, somente após esta carga ser assimilada, se sobrecarrega a intensidade (DANTAS, 2014).

4) O princípio da interdependência volume/intensidade trata da relação inversa entre volume e intensidade no contexto do treinamento esportivo. Quando realiza-se uma atividade muito intensa, não é possível manter tal intensidade por um longo período de tempo, enquanto em uma atividade com menor intensidade consegue-se manter durante um período mais prolongado. Durante um ciclo de treinamento, o volume será enfatizado em determinada fase, como períodos de transição, e a intensidade será enfatizada em outras fases, como períodos competitivos (DANTAS, 2014).

5) O princípio da continuidade diz respeito à manutenção da rotina de treinamento, sem interrupções, respeitando-se os períodos de recuperação. A interrupção controlada, para fins de recuperação, é benéfica e essencial para alcançar os objetivos desejados, geralmente variando de poucos minutos até 40 horas. Após 48 horas pode iniciar um processo de perda de condicionamento físico caso não ocorra um novo estímulo, ou seja, a continuidade do treinamento. De modo geral, pausas de mais de 48 horas são necessárias somente em casos de

11 Homeostase é o equilíbrio do ambiente interno do corpo humano por meio de respostas fisiológicas, mantido apesar de mudanças no ambiente externo (BARBANTI, 2003).

*overtraining*¹² ou, obviamente, em caso de lesão. Um período de interrupção igual ou superior a 4 semanas é suficiente para perder todo o trabalho realizado em termos de condicionamento físico (DANTAS, 2014).

6) O princípio da especificidade trata do foco específico do treinamento em função da qualidade física que se deseja alcançar (velocidade, resistência anaeróbica ou resistência aeróbica). O treinamento deverá ser direcionado de acordo com os requisitos específicos, como sistema energético preponderante e segmentos corporais utilizados. O planejamento do treinamento deverá levar em conta, com precisão, a via energética preponderante e, conforme a via utilizada, o treinamento priorizará determinada qualidade física (DANTAS, 2014). O princípio da especificidade implica que determinados tipos de treinamento sejam mais adequados do que outros para desenvolver componentes específicos da aptidão física (HEYWARD, 2013).

Além dos princípios básicos do treinamento esportivo, Heyward (2013) destaca alguns elementos primordiais para a prescrição de exercícios, tais como a intensidade, a duração (volume) e a frequência. A intensidade do exercício determina mudanças fisiológicas e metabólicas específicas no organismo durante o exercício físico e depende de diversos fatores, tais como objetivo desejado, idade e nível de aptidão física. O volume é inversamente proporcional à intensidade. O volume das atividades dependerá também do estado de saúde do praticante, nível de aptidão física, capacidade funcional e objetivos. Tanto o volume quanto a intensidade podem ser gradualmente aumentados conforme ocorre adaptação ao treinamento. A frequência geralmente se refere à quantidade semanal de sessões, mas também está diretamente relacionada ao volume e à intensidade, pois será determinada essencialmente pela necessidade de períodos maiores ou menores de recuperação.

12 Overtraining, no contexto do treinamento esportivo, diz respeito ao excesso de treinamento, sem respeitar um tempo adequado de recuperação, podendo resultar em prejuízos como maior incidência de lesões e diminuição do rendimento (BARBANTI, 2003).

3.2.3 Saúde e qualidade de vida

Saúde e qualidade de vida são conceitos que, de maneira geral, se complementam. O processo natural de envelhecimento, caracterizado por mudanças fisiológicas no corpo humano, pode afetar nossa capacidade de realizar determinadas tarefas, impactando diretamente na qualidade de vida. Muitos estudos (BIZE; JOHNSON; PLOTNIKOFF, 2007; FOX et al., 2007; VAGETTI et al., 2015; WENDEL-VOS et al., 2004) apresentam os benefícios da prática regular de atividade física para a qualidade de vida em diversas faixas etárias.

Segundo o Grupo de Qualidade de Vida da Organização Mundial da Saúde (OMS) (WORLD HEALTH ORGANIZATION, 1998), qualidade de vida é definida como “a percepção do indivíduo de sua posição em sua vida, em um contexto cultural e sistemas de valores nos quais ele vive, além da relação a seus objetivos, expectativas, padrões e preocupações”. Diante da necessidade de um instrumento para avaliação da percepção da qualidade de vida em um contexto transcultural, inicialmente foi desenvolvido o instrumento *World Health Organization Quality of Life* (WHOQOL-100), contendo 100 perguntas, envolvendo a participação de diversos países e representantes de diferentes culturas (FLECK et al., 2000).

Um questionário com 100 perguntas pode ser inviável devido à dificuldade de ser aplicado em diversas situações. Dessa forma, com a necessidade de instrumentos curtos, sem deixar de preservar características psicométricas satisfatórias, o Grupo de Qualidade de Vida da OMS criou o instrumento WHOQOL Abreviado (WHOQOL-bref), uma versão abreviada do WHOQOL-100 contendo 26 perguntas, sendo duas delas sobre percepção da qualidade de vida de um modo geral e as demais representando cada uma das 24 facetas que compõem o instrumento original, conforme o Quadro 1 (FLECK et al., 2000).

Quadro 1 – Domínios e facetas do instrumento WHOQOL Abreviado

Domínios	Facetas
1 – Domínio Físico	1 – Dor e desconforto 2 – Energia e fadiga 3 – Sono e repouso 9 – Mobilidade 10 – Atividades da vida cotidiana 11 – Dependência de medicação e tratamentos 12 – Capacidade de Trabalho
2 – Domínio psicológico	4 – Sentimentos positivos 5 – Pensar, aprender, memória e concentração 6 – Autoestima 7 – Imagem corporal e aparência 8 – Sentimentos negativos 24 – Espiritualidade, religião, crenças pessoais
3 – Relações sociais	13 – Relações pessoais 14 – Suporte (apoio) socialização 15 – Atividade sexual
4 – Meio ambiente	16 – Segurança física e proteção 17 – Ambiente no lar 18 – Recursos financeiros 19 – Cuidados de saúde e sociais: qualidade e disponibilidade 20 – Oportunidades de adquirir novas informações e habilidades 21 – Oportunidades de recreação e lazer 22 – Ambiente físico (poluição, ruído, trânsito, clima) 23 – Transporte

Fonte: Adaptado de Fleck et al. (2000)

Diversos autores apresentam os benefícios da atividade física para a qualidade de vida. Culos-Reed e Brawley (2000) estudaram dois grupos, um praticava atividade física até duas vezes por semana e outro praticava pelo menos três vezes por semana. Os indivíduos que praticavam atividade física pelo menos três vezes por semana apresentaram maiores scores de qualidade de vida no componente físico. Stewart et al. (2003) avaliaram três níveis de atividade física em relação à qualidade de vida: moderada, vigorosa e muito vigorosa. De acordo com seus resultados, quanto maior o nível de atividade física maior o score de qualidade de vida. Mummery, Schofield e Caperchione (2004) compararam três grupos: inativos, moderadamente ativos e ativos. Os grupos “moderadamente ativos” e “ativos” apresentaram scores significativamente maiores nos componentes de qualidade de vida físico e mental, quando comparados aos inativos. Pucci et al.

(2012a) estudaram a relação entre qualidade de vida e atividade física em 1461 adultos brasileiros, analisando a frequência semanal, e concluíram que os indivíduos mais ativos (mais de 150 minutos por semana) apresentam maiores *scores* em todos os domínios do instrumento de avaliação da qualidade de vida. Em uma revisão de 38 estudos relacionados à associação entre atividade física e qualidade de vida em adultos publicados a partir de 2003, Pucci et al. (2012b) identificaram que um maior nível de atividade física está associado a melhor percepção da qualidade de vida.

Embora os benefícios da atividade física sejam amplamente relatados na literatura, não podemos esquecer dos riscos envolvidos. A prática de atividade física em qualquer nível de performance envolve riscos relacionados a fatores físicos e psicológicos. A participação em atividades esportivas envolve uma elevada demanda física, exigindo ajustes no organismo que podem resultar em efeitos negativos no corpo humano, devido ao potencial para desequilíbrio musculoesquelético, podendo resultar em lesões. Tais lesões podem afetar o atleta tanto fisicamente quanto psicologicamente, levando a, por exemplo, a diminuição de performance e afastamento de competições, afetando diretamente a qualidade de vida (MCINTOSH, 2005).

3.2.4 Influência do clima na prática de atividade física

A atividade física pode ser praticada nos mais diversos ambientes. Quando se trata de atividades ao ar livre, podem ser realizadas nos mais variados locais, tais como parques, trilhas, ruas, rodovias e montanhas. A variedade de locais disponíveis traz consigo diferentes condições climáticas, tais como frio, calor, chuva, baixa ou alta umidade do ar, entre outras. Tais condições climáticas exercem um impacto direto no corpo humano, podendo afetar de alguma forma o desempenho, seja qual for a modalidade praticada. Dessa forma, é de fundamental importância a manutenção, por exemplo, de um equilíbrio térmico durante a prática de atividade física. Para lidar com as variações climáticas, o corpo humano conta com diversos mecanismos de regulação com o objetivo de manter a homeostase (POWERS; HOWLEY, 2000).

Segundo Powers e Howley (2000), a prática de atividade física em um ambiente quente e úmido é um desafio para a manutenção da temperatura corporal, pois o calor elevado combinado com a umidade elevada reduzem a capacidade de perda de calor, o que resulta em uma maior taxa de sudorese. A combinação de perda líquida por meio da sudorese e o aumento da temperatura central aumentam o risco de hipertermia. O oposto ocorre com a prática de atividade física em um ambiente frio, aumentando a capacidade de perda de calor. A produção de calor natural por meio do metabolismo, combinada com a utilização de roupas quentes adequadas para o ambiente frio, em geral impedem a hipotermia. No entanto, atividades prolongadas em clima frio podem extrapolar a capacidade do corpo humano de evitar a perda de calor, principalmente atividades que envolvam contato com a água ou em condições de chuva, e, dessa forma, levar a um quadro de hipotermia.

Tanto a hipertermia quanto a hipotermia são exemplos de fatores frequentemente presentes em decorrência da prática de atividade física em condições climáticas extremas, e podem ser evitados, ou, pelo menos minimizados, por meio de um planejamento adequado. A temperatura do corpo humano, em condições normais, deve variar entre 35º C e 37,5º C, de modo que a exposição a ambientes muito frios aumenta o risco de hipotermia, quando a temperatura corporal diminui abaixo de 35º C, bem como a exposição a ambientes muito quentes aumenta o risco de hipertermia, quando a temperatura corporal aumenta acima de 37,5º C (PRESSLER; NIEBAUER, 2020).

Pressler e Niebauer (2020) apresentam algumas formas de prevenir ou amenizar problemas relacionados à prática de atividade física em condições climáticas extremas: o atleta deve ter um nível de aptidão física adequado para os desafios específicos a serem enfrentados; realizar uma aclimatação prévia, ou seja, um treinamento que simule as condições a serem enfrentadas; realizar testes e consultas médicas durante a preparação; utilizar roupas e equipamentos adequados.

Diversos autores (HÖPPE, 1999; MATZARAKIS; MAYER; IZIOMON, 1999; PALLOTTA; HERDIES; GONÇALVES, 2015; SUPING et al., 1992) apresentam

índices biometeorológicos ou índices de conforto térmico, calculados por meio de equações empíricas, com o objetivo de estimar o quanto determinadas variáveis meteorológicas influenciam parâmetros fisiológicos do corpo humano. Segundo Pezzoli et al. (2012), a avaliação das condições bioclimatológicas e de conforto térmico em competições esportivas, como ciclismo e atletismo, é de fundamental importância para o planejamento do programa de treinamento e definição de estratégias para a competição. No entanto, a influência das condições meteorológicas e ambiental são, geralmente, negligenciadas na avaliação de desempenho em esportes ao ar livre.

3.3 Tecnologia e atividade física

Inúmeras tecnologias, basicamente mecânicas ou eletromecânicas, estão presentes em equipamentos voltados para a prática de atividade física, como aparelhos de musculação, esteiras rolantes e bicicletas estacionárias. Com a popularização e redução de custos de componentes eletrônicos, algumas empresas investem no aprimoramento dos equipamentos por meio da integração com dispositivos móveis utilizando tecnologias de conexão sem fio como Bluetooth e NFC¹³. A Figura 3 mostra a possibilidade de conexão do relógio *Apple Watch* com uma esteira rolante por meio da tecnologia NFC.

13 Sigla para *Near-field Communication*, um protocolo de comunicação de dados entre dispositivos eletrônicos (CAMERON, 2017).

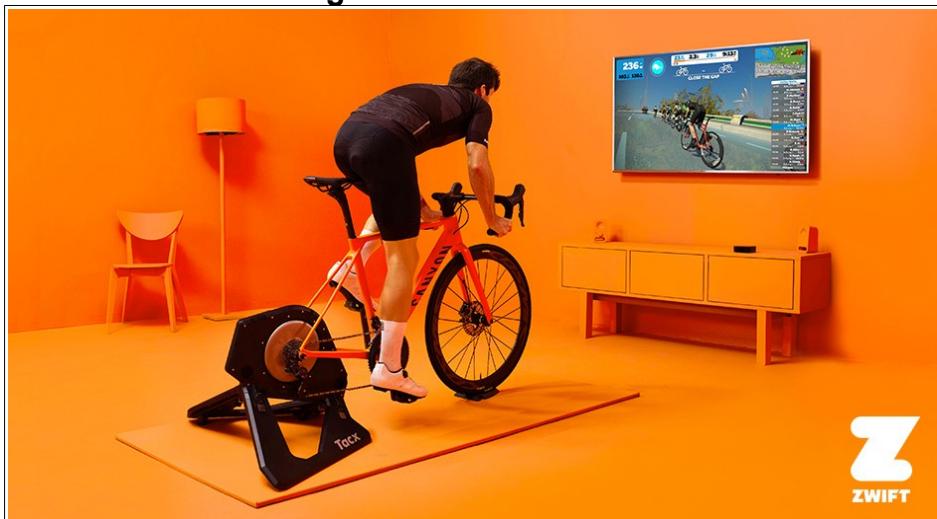
Figura 3 – Integração entre aparelho de academia e Apple Watch



Fonte: Pure Fitness (PURE FITNESS, 2017)

A tecnologia tem sido também empregada em grande escala em roupas e acessórios esportivos (YANFEN; PU, 2011), além de produtos de entretenimento voltados para a prática de atividade física, principalmente na forma de jogos simuladores, como o *Zwift*¹⁴ (Figura 4).

Figura 4 – Sistema Zwift



Fonte: Castelli Insider News (CASTELLI, 2020)

Fiordelli, Diviani e Schulz (2013) destacam a crescente expansão no início da década de 2010 do conceito “*mobile health*” (*mHealth*), definido como o uso de

14 Disponível em <https://www.zwift.com/>

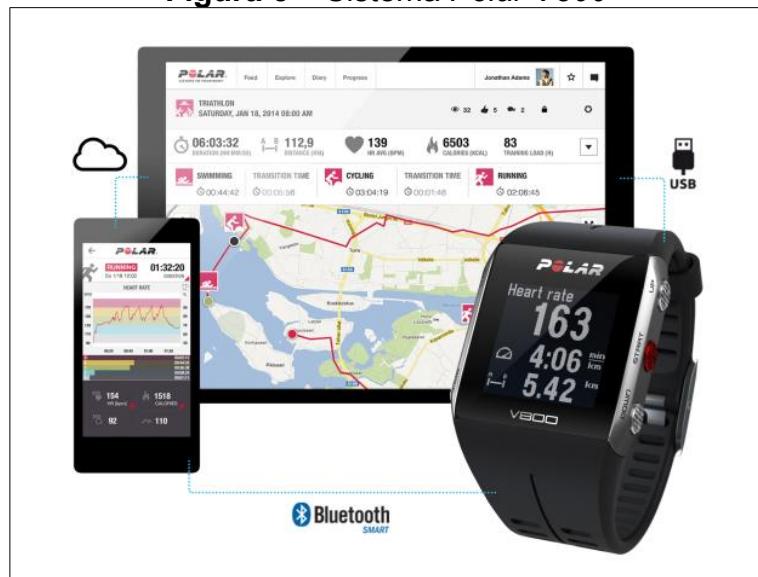
computação móvel e tecnologias de comunicação aplicadas à saúde. As aplicações móveis voltadas para a saúde de modo geral abrangem diversos tipos de usuários, tais como profissionais, pacientes ou mesmo pessoas saudáveis. Os *smartphones* oferecem cada vez mais recursos voltados para diferentes propostas, tais como emagrecimento e monitoramento de atividade física, tendo como grande vantagem o fato de estarem sempre conectados à Internet.

Quando pensamos na produção de dados, não se trata somente daqueles obtidos por meio de dispositivos de entrada em computadores, como teclados ou telas sensíveis ao toque. Além da tradicional entrada de forma manual (por exemplo, planilhas), os dados podem ser produzidos por meio do processamento ou transformação de outros dados existentes (AMARAL, 2016). Ainda, com o tamanho cada vez mais reduzido e custo cada vez mais baixo, sensores dos mais diversos tipos estão cada vez mais presentes em *smartphones*, relógios e pulseiras, além de muitos outros equipamentos. Tais sensores possibilitam a obtenção dos mais diversos tipos de registros, especialmente no que diz respeito ao monitoramento de atividade física e saúde, tais como coordenadas de localização por GPS, frequência cardíaca, velocidade, distância e altimetria.

Atualmente, o mercado de consumo oferece uma infinidade de dispositivos voltados para o monitoramento da prática de atividade física, desde simples pulseiras capazes de apenas estimar uma contagem de passos até relógios contendo as mais recentes tecnologias embarcadas, capazes de medir a variabilidade da frequência cardíaca, como o modelo V800 (Figura 5) da marca *Polar* (GILES; DRAPER; NEIL, 2016), ou mesmo realizar um eletrocardiograma (ECG), como o *Apple Watch* (Figura 6) a partir da quarta geração (AVILA, 2019). Tais equipamentos contam com diversos sensores que, utilizados de forma individual ou combinada, proporcionam a obtenção de diversos dados. Por meio de uma antena GPS, é possível medir a distância percorrida, a altimetria e, combinando o deslocamento geográfico com o cronômetro, a velocidade em tempo real. Acelerômetros possibilitam estimar a quantidade de passos durante uma caminhada. Com uma fita contendo eletrodos instalada no peito do atleta, ou um dispositivo de

pulso baseado em fotopletismografia¹⁵ (KOSHY et al., 2018), é possível aferir a frequência cardíaca tanto em repouso quanto durante a prática de atividade física.

Figura 5 – Sistema Polar V800



Fonte: Canal de Bike (OLIVEIRA, 2015)

15 Fotopletismografia é um método óptico, não invasivo e de baixo custo, utilizado para detectar variações no volume sanguíneo em tecidos microvasculares, sendo a medida realizada na superfície da pele (BISWAS et al., 2019)

Figura 6 – Apple Watch com ECG



Fonte: *Apple Inc. (APPLE INC, 2020)*

Apesar da grande quantidade de aplicativos disponíveis para o monitoramento de atividade física, capazes de coletar dados, e utilizados diariamente por milhões de praticantes, ao verificar as funcionalidades daqueles mais populares (Figura 7), constata-se que não existem, até o momento, mecanismos capazes de oferecer, principalmente ao profissional de Educação Física, de forma simples e acessível, por meio da Mineração de Dados, *feedback* e sugestões que possam auxiliar na tomada de decisões para uma melhor qualidade da prescrição e acompanhamento do treinamento.

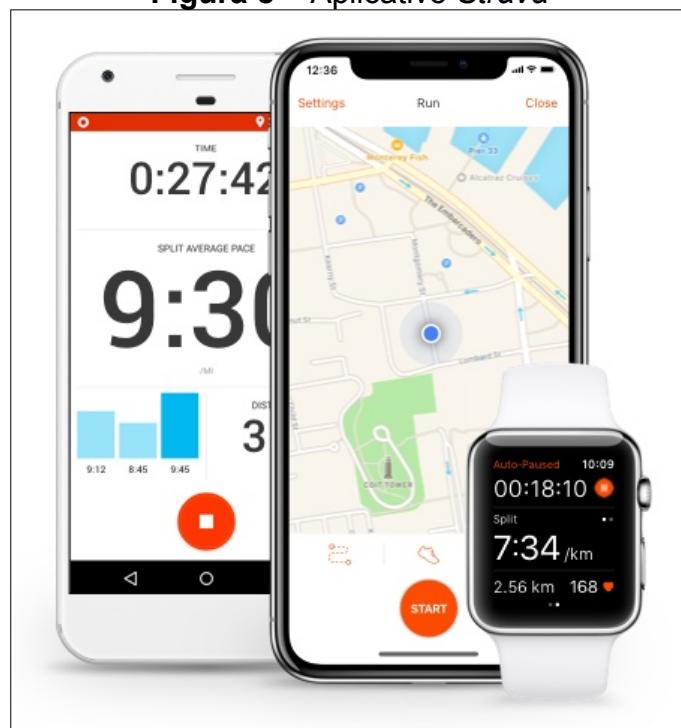
O aplicativo *Strava* (Figura 8) foi criado para que ciclistas e corredores de rua possam registrar suas atividades e analisar a performance e se tornou bastante popular devido a uma ênfase competitiva, por meio da comparação de desempenho em determinadas rotas ou determinados segmentos com outros usuários (WEST, 2015).

Figura 7 – Aplicativos mais populares na categoria “Saúde e Atividade Física” em Fevereiro de 2020



Fonte: SensorTower (WILLIAMS, 2020)

Figura 8 – Aplicativo Strava



Fonte: Strava (2020a)

O *Strava* alcançou a marca de 50 milhões de usuários em fevereiro de 2020, com 3 bilhões de atividades registradas desde sua criação em 2009 (STRAVA, 2020b). O aplicativo é gratuito e um dos mais populares na categoria relacionada à saúde e atividade física, ocupando uma das primeiras posições no ranking de aplicativos mais baixados (Figura 7) nas duas principais plataformas de distribuição de aplicativos móveis (*Apple App Store* para dispositivos com sistema *Apple iOS* e *Google Play Store* para dispositivos com sistema *Android*).

A utilização de uma API de acesso livre e que possibilite recuperar dados referentes a uma grande quantidade de variáveis, de uma quantidade expressiva de participantes, foi fundamental para viabilizar a coleta de dados para o desenvolvimento do presente estudo. O próximo capítulo apresenta, de forma detalhada, os procedimentos metodológicos adotados para o desenvolvimento da pesquisa.

4 MATERIAL E MÉTODOS

Este capítulo detalha os procedimentos metodológicos adotados, apresentando informações pertinentes à amostra utilizada, variáveis, etapas de coleta e armazenamento dos dados, tratamento e análise dos dados.

4.1 Amostra

A amostra é composta por 1059 participantes, usuários do aplicativo *Strava*. A Tabela 1 resume as características da amostra. Para o presente estudo, optou-se por selecionar somente as atividades de ciclismo e corrida de rua, por se tratarem das atividades mais populares, correspondendo a maior parte das atividades.

Tabela 1 – Caracterização da amostra

Sexo	N	Idade (anos)	Massa Corporal (kg)	Estatura (cm)	IMC (kg/m ²)
Masculino	844	38 ± 11	78,2 ± 16,6	176 ± 8	25,2 ± 5,6
Feminino	215	38 ± 10	62,9 ± 20,5	163 ± 7	24,4 ± 6,1

Fonte: Elaborado pelo autor

O tamanho mínimo recomendado para a amostra foi calculado conforme a equação (1) para populações finitas (LEVINE; BERENSON; STEPHAN, 2014), considerando a quantidade de usuários registrados no aplicativo *Strava* (50 milhões até fevereiro de 2020), sendo “n” o tamanho da amostra a ser calculado, “N” o tamanho da população, “σ” a proporção populacional de indivíduos que pertence à categoria de interesse, “Z_{α/2}” o valor crítico relacionado ao grau de confiança (Quadro 2) e “E” representa a margem de erro.

$$n = \frac{N \cdot \sigma^2 \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot E^2 + \sigma^2 \cdot (Z_{\alpha/2})^2} \quad (1)$$

Quadro 2 – Valores críticos associados ao grau de confiança na amostra

Grau de confiança	α	$Z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

Fonte:Levine, Berenson e Stephan (2014)

Para o presente estudo foi utilizado grau de confiança de 99% ($Z_{\alpha/2}=2,575$) e margem de erro de 5% ($E=0,05$). Como a proporção populacional de indivíduos que pertence à categoria de interesse para o presente estudo é desconhecida, foi utilizado 50% como valor padrão ($\sigma=0,5$), conforme sugerido por Levine, Berenson e Stephan (2014). De acordo com o resultado do cálculo, a amostra deveria ser composta por, pelo menos, 663 participantes.

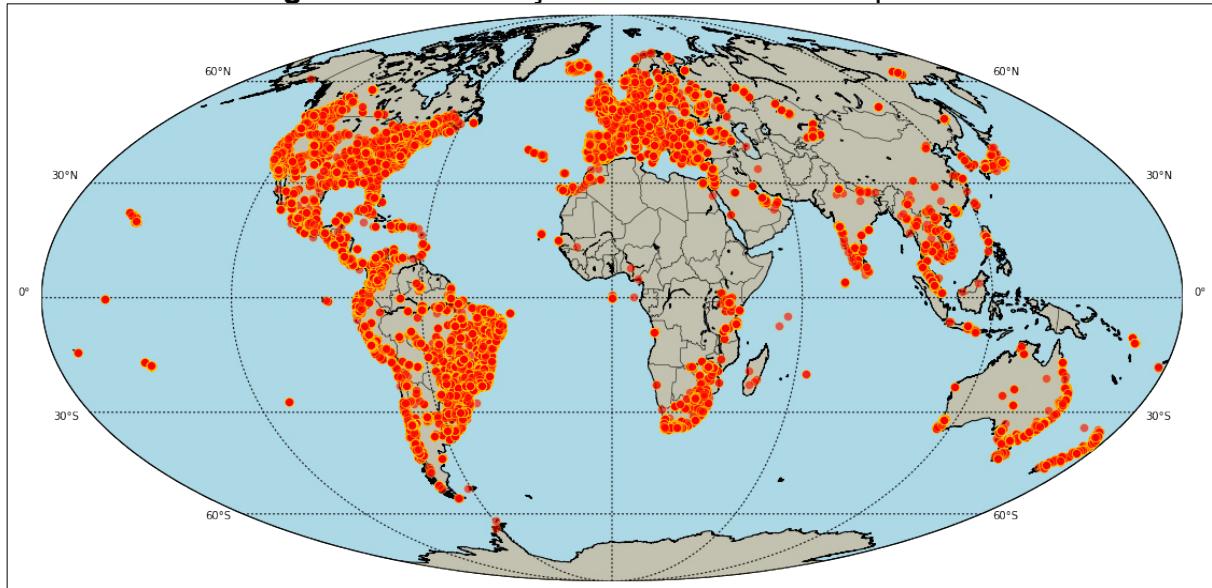
Para a seleção da amostra, foram adotados os seguintes critérios de inclusão:

- O participante deveria ter idade entre 18 e 65 anos;
- O participante deveria ser usuário do aplicativo *Strava*, ter registrado atividades de ciclismo ou corrida de rua e autorizar o compartilhamento de seus dados;
- As atividades deveriam ter no mínimo 30 minutos de duração.

Além do compartilhamento dos dados do *Strava*, todos os participantes responderam ao questionário WHOQOL Abreviado (Anexo C) (FLECK et al., 2000), em formato online, disponibilizado na forma de uma página da Internet compondo o sistema de coleta de dados.

Os participantes realizaram, no total, 503.369 atividades, em 136 países. A Figura 9 apresenta a distribuição das atividades coletadas (pontos vermelhos na figura).

Figura 9 – Distribuição da amostra em 136 países



Fonte: Elaborado pelo autor

O Projeto de Pesquisa para a realização do presente estudo foi aprovado pelo Comitê de Ética em Pesquisa com seres humanos do Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP, campus de Rio Claro/SP, conforme parecer nº 3.064.004 (Anexo A), e registrado na Plataforma Brasil conforme protocolo CAAE nº 93316318.6.0000.5465.

Quanto aos riscos relacionados à ética em pesquisa, a utilização de questionário online é sensível no que diz respeito à invasão de privacidade e vazamento de informações confidenciais do participante. Tais riscos foram tratados, de forma a minimizá-los, por meio de medidas visando o anonimato do participante, sendo a identidade do participante codificada por meio de chave numérica, além de medidas de proteção aos dados, tais como criptografia tanto no armazenamento quanto no transporte dos dados, por meio da tecnologia *Secure Socket Layer* (SSL), um padrão global de segurança que cria um canal criptografado entre o servidor *web*¹⁶ e o *software* navegador de Internet do usuário, proporcionando sigilo e segurança na transmissão dos dados, além de *firewall*¹⁷ e outras medidas de

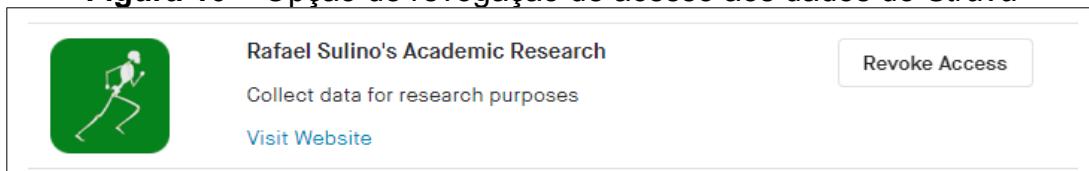
16 Servidor *web* é um *software* de servidor que utiliza o protocolo HTTP para apresentar documentos HTML (páginas da Internet) e outros arquivos associados quando solicitados pelo cliente (por exemplo, navegador *web*) (MICROSOFT CORPORATION, 1998).

17 *Firewall* é um tipo de sistema de segurança de rede que monitora o tráfego de entrada e saída e decide permitir ou bloquear tráfegos específicos, de acordo com um conjunto definido de regras de segurança (COLE; KRUTZ; CONLEY, 2005).

segurança oferecidas pela empresa que hospeda o servidor (conforme descrito no item 4.2). Ainda, não foi registrado o endereço IP do participante, de modo a reforçar os procedimentos que visam o anonimato. Quanto ao Termo de Consentimento Livre e Esclarecido – TCLE (Apêndice A), este foi disponibilizado aos participantes em formato virtual (vide Decisão CEP 01/2019, Anexo B), digitalizado, nos idiomas português, inglês e espanhol, assinado pelo pesquisador responsável por meio de certificado digital em conformidade com o padrão da Infraestrutura de Chaves Públicas Brasileiras (ICP-Brasil), conforme a Lei nº 11.419/2016. Cabe ressaltar que a própria característica de anonimato da pesquisa não contempla a devolução do TCLE assinado pelo participante. Dessa forma o TCLE tradicional foi substituído pelo TCLE virtual, apresentado na forma de convite, com campo onde o participante possa manifestar seu aceite ou não em participar da pesquisa, em conformidade com o item 8.2 da Decisão do CEP/IB 01/2019 (Anexo B).

De modo a atender ao item 9.3 da Decisão do CEP/IB 01/2019 (Anexo B), o participante pode cancelar sua participação na pesquisa a qualquer momento, revogando o acesso por meio de sua conta do aplicativo *Strava*, conforme informado no momento da autorização. Para tanto, o participante deve acessar as configurações em sua conta e, em sua lista de aplicativos conectados, efetuar a revogação do acesso a seus dados (Figura 10).

Figura 10 – Opção de revogação de acesso aos dados do *Strava*



Fonte: Elaborado pelo autor

Por meio de uma revalidação dos cadastros, realizada diariamente pelo servidor, um algoritmo identifica o registro revogado, conforme resposta da API do *Strava* para requisições HTTP¹⁸, utilizando o código numérico identificador do participante e a chave (*token*) de autorização. Na revalidação dos cadastros, é

¹⁸ HTTP é a sigla de “*Hypertext Transfer Protocol*” (protocolo de transferência de hipertexto). Trata-se de um protocolo de transferência de informações para a comunicação de dados da Internet (MICROSOFT CORPORATION, 1998).

esperado o código de resposta HTTP “200” (autorizado), de forma a confirmar a autorização. No caso de resposta HTTP com código “401” (não autorizado), o algoritmo procede com a exclusão de quaisquer dados associados ao código do participante desistente. Até o momento, não houve nenhum registro de cancelamento de autorização.

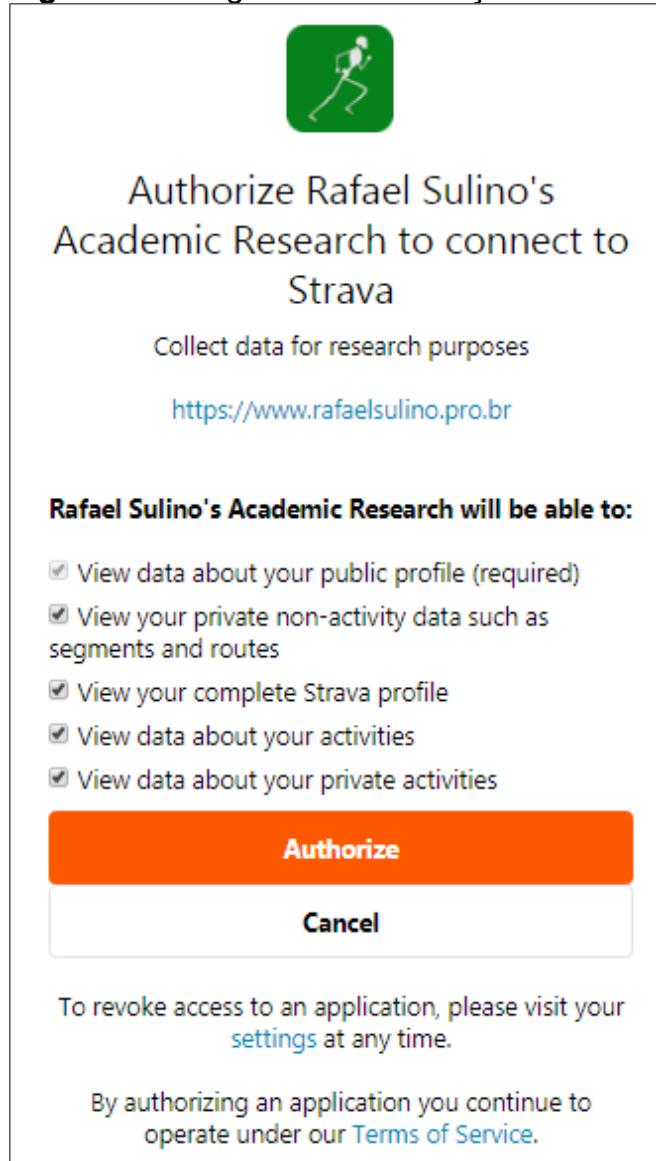
4.2 Coleta e armazenamento de dados

A coleta de dados foi realizada de forma online, no período entre 15 de setembro de 2019 e 15 de abril de 2020. Em um primeiro momento, foi realizada a divulgação por meio do envio de convites (Apêndices D, E e F), em três idiomas (português, inglês e espanhol), para 350 grupos cadastrados na rede social interna do aplicativo *Strava*, e também por meio da rede social *Instagram*¹⁹ (Apêndices G e H). Todos os elementos gráficos envolvendo a marca da Universidade Estadual Paulista “Júlio de Mesquita Filho” - Unesp foram utilizados em conformidade com o manual de uso (ROSA; SILVEIRA, 2019).

Ao acessar o endereço eletrônico da pesquisa, o participante teve acesso a uma carta convite explicando a pesquisa, bem como ao *link* para *download* do TCLE virtual contendo todas as informações pertinentes. Após concordar com a participação, o participante foi direcionado para a página de autenticação (Figura 11) do *Strava*, na qual forneceu permissão para compartilhamento de seus dados. Em seguida, foi direcionado para a página do questionário WHOQOL Abreviado (Apêndice C), etapa final da coleta. Todas as etapas da coleta de dados foram projetadas com atenção especial à característica de anonimato do participante.

¹⁹ Disponível em <https://www.instagram.com>

Figura 11 – Página de autenticação do Strava



Fonte: Elaborado pelo autor

Para a coleta, armazenamento e processamento de dados, foi utilizado um Servidor Virtual Privado (VPS), com sistema operacional *Linux Debian*, versão 9.11 para arquitetura computacional de 64 bits, 200 gigabytes de espaço de armazenamento e 64 gigabytes de memória RAM. Os recursos de hardware deste tipo de servidor são escaláveis, de forma que podem ser ajustados conforme a demanda do projeto. A Figura 12 resume as configurações do servidor usado.

Figura 12 – Configurações do servidor VPS

```
Last login: Mon Jan 20 17:58:58 2020 from 170.0.207.237
      ,met$$$$$gg.          root@rafaelsulino.pro.br
      ,g$$$$$$$$$$$$$$$$$P.  -----
      ,$$P"      """"Y$$. .
      ,$$P'      ` $$$.      OS: Debian GNU/Linux 9.11 (stretch) x86_64
      ,$$P      ,ggs.      ` $$b:  Kernel: 4.15.0-74-generic
      d$$'      ,$$P"'.     $$$  Uptime: 8 hours, 43 minutes
      $$P      d$'      ,   $$P  Packages: 1067
      $$:      $$.      - ,d$$'  Shell: bash 4.4.12
      $$;      Y$b.      ,d$P'  CPU: Intel Xeon E-2136 (2) @ 4.5GHz
      Y$$.      . "Y$$$P"  Memory: 1251MB / 64208MB
      $$b      "-.-
      `Y$$
      `Y$$.      .
      `$$b.
      `Y$$b.
      `Y$b.      .
      `Y$b.      .
      `Y$b.      .

(base) root@rafaelsulino:~# |
```

Fonte: Elaborado pelo autor

Quanto à infraestrutura que abriga o servidor VPS (*datacenter*), ela está localizada geograficamente em Montreal no Canadá.

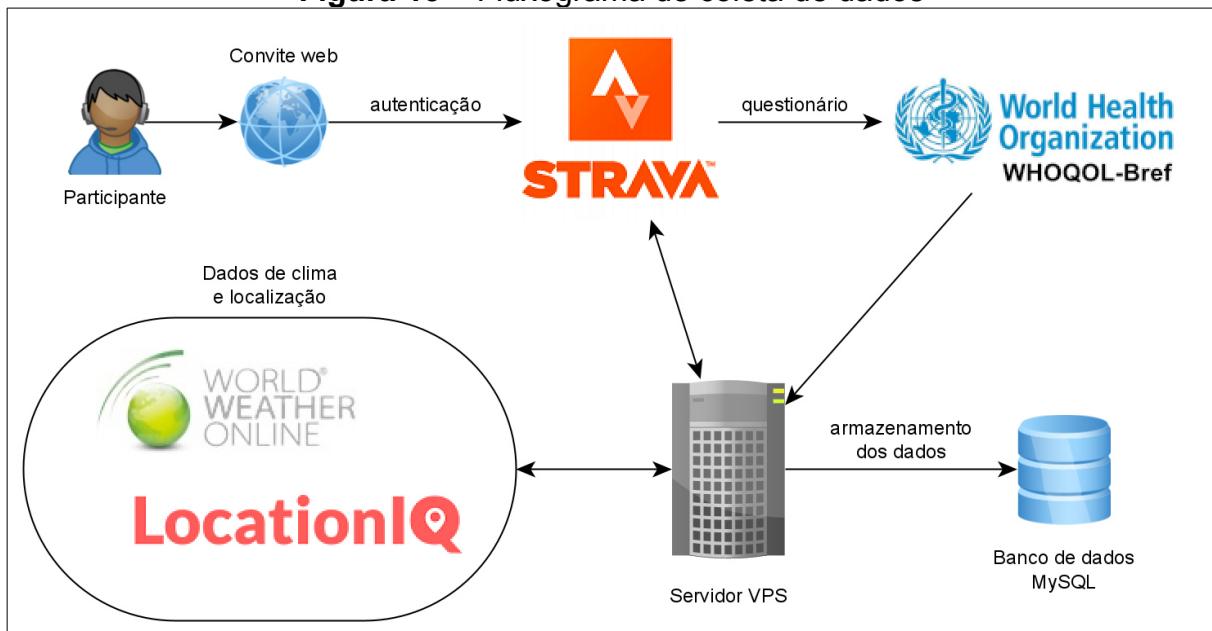
A aquisição de dados climáticos, tais como temperatura, umidade do ar, velocidade e direção do vento, pressão atmosférica, fenômenos como chuva e neblina, entre outras variáveis, foi realizada por meio de acesso à API do serviço *World Weather Online*, usando data, horário e localização geográfica (coordenadas GPS) das atividades praticadas.

O sistema de aquisição de dados (Figura 13) foi composto por duas partes, denominadas *frontend* e *backend*. O *frontend* corresponde às páginas na Internet a serem acessadas pelos participantes, para que possam autorizar o compartilhamento de seus dados. Ele é composto por uma página de apresentação, um breve questionário com questões relacionadas à saúde e atividade física, o questionário WHOQOL Abreviado (Apêndice C) e autenticação para autorização de compartilhamento dos dados do aplicativo *Strava*. Já o *backend* corresponde a APIs no servidor, desenvolvidas em *Node.js*²⁰, uma plataforma de desenvolvimento de aplicações em servidor baseada na linguagem *Java Script*, que é responsável por

20 Disponível em <https://nodejs.org>

buscar os dados de atividade física de cada usuário, buscar dados climáticos referentes a cada atividade registrada e armazenar ou recuperar informações no banco de dados. Os dados são armazenados em um banco de dados MySQL Community Server²¹, versão 5.7.26.

Figura 13 – Fluxograma de coleta de dados



Fonte: Elaborado pelo autor

Com o objetivo de prevenir falhas e perda de dados, o servidor é constantemente monitorado pelo serviço gratuito *Uptime Robot*²², o qual envia alertas ao administrador do servidor ao detectar falhas e indisponibilidade, possibilitando uma imediata intervenção no sentido de regularizar seu funcionamento. O servidor conta ainda com a realização de *backup* automático diariamente, utilizando o serviço de armazenamento de arquivos *Google Drive*²³.

Alguns dos aplicativos e dispositivos para monitoramento de atividade física disponíveis no mercado oferecem uma interface de programação (API), por meio da qual é possível obter os dados via acesso remoto e utilizá-los para análises com ferramentas de estatística e linguagens de programação de computadores. As

21 Disponível em <https://dev.mysql.com>

22 Disponível em <https://uptimerobot.com>

23 Disponível em <https://drive.google.com>

opções cogitadas inicialmente para o presente estudo foram: *Strava API*, *Fitbit SDK*²⁴, *Garmin Health API*²⁵, *Polar Open Accesslink*²⁶ e *Under Armour Connected Fitness*²⁷. As APIs *Garmin* e *Polar* foram descartadas por não serem de acesso livre. A API *Fitbit* foi descartada por não permitir em sua rede social interna a divulgação para fins de pesquisa, o que dificultaria o recrutamento de participantes. Restando as APIs *Strava* e *Under Armour*, a *Strava* foi escolhida por ser de acesso gratuito, ter milhões de usuários registrados e contar com uma rede social interna bastante ativa, a qual viabilizou a divulgação e recrutamento de participantes. Outro fator determinante para a escolha do *Strava* diz respeito ao limite diário de requisições de acesso, pois oferece um limite de 600 requisições a cada 15 minutos, totalizando 30 mil requisições por dia, enquanto outras plataformas, de modo geral, limitam a 500 requisições por dia ou cobram uma espécie de assinatura para que se tenha uma quantidade maior ou ilimitada de requisições.

4.3 Variáveis coletadas

Os dados coletados correspondem a diversas variáveis, de diferentes fontes. As variáveis relacionadas à prática de atividade física foram obtidas por meio da API do *Strava*, sendo que os dados que envolvem frequência cardíaca foram obtidos somente daqueles que utilizaram dispositivo específico para tal finalidade. As variáveis climáticas foram obtidas por meio da API do serviço *World Weather Online*²⁸. Ainda, foram obtidas variáveis referentes à localização, com o objetivo de identificar a cidade e país onde a atividade foi realizada, por meio da API do serviço *LocationIQ*²⁹. As variáveis relacionadas à qualidade de vida foram obtidas por meio do preenchimento online do instrumento WHOQOL Abreviado.

24 Disponível em <https://dev.fitbit.com>

25 Disponível em <https://developer.garmin.com>

26 Disponível em <https://www.polar.com/accesslink-api>

27 Disponível em <https://developer.underarmour.com>

28 Disponível em <https://www.worldweatheronline.com>

29 Disponível em <https://locationiq.com>

O Quadro 3 apresenta as variáveis obtidas por meio da API do *Strava*.

Quadro 3 – Descrição das variáveis obtidas por meio da API do *Strava*

Variável	Descrição
<code>id_participante</code>	Código identificador único de cada participante
<code>id_atividade</code>	Código identificador único de cada atividade
<code>tipo</code>	Tipo de atividade (por exemplo, corrida, ciclismo, natação)
<code>data</code>	Data da atividade
<code>hora_inicio</code>	Hora em que a atividade foi iniciada
<code>latitude_inicio</code>	Geolocalização
<code>longitude_inicio</code>	Geolocalização
<code>distance</code>	Distância em metros
<code>moving_time</code>	Tempo em movimento (sem pausa)
<code>elapsed_time</code>	Tempo total (incluindo pausas)
<code>total_elevation_gain</code>	Ganho de elevação acumulado, em metros
<code>max_speed</code>	Velocidade máxima em m/s
<code>average_speed</code>	Velocidade média em m/s
<code>average_cadence</code>	Cadência de pedalada (específico ciclismo)
<code>average_watts</code>	Potência média de pedaladas em watts (específico ciclismo)
<code>kilojoules</code>	Dispêndio de energia
<code>calories</code>	Convertido joules em Kcal
<code>average_heartrate</code>	Frequência cardíaca média da atividade
<code>max_heartrate</code>	Frequência cardíaca máxima obtida na atividade
<code>suffer_score</code>	Medida de esforço baseada nas zonas de treinamento e o tempo de permanência em cada zona (zonas mais intensas têm maior peso, ou seja, quanto maior o <i>suffer score</i> , maior o tempo de permanência em intensidade elevada)
<code>elev_high</code>	Altitude máxima do percurso
<code>elev_low</code>	Altitude mínima do percurso
<code>avg_speed_kph</code>	Velocidade média em km/h
<code>max_speed_kph</code>	Velocidade máxima em km/h
<code>tempo_pausa</code>	Soma das pausas ocorridas durante a atividade

Fonte: Elaborado pelo autor

O Quadro 4 apresenta as variáveis obtidas por meio da API do serviço *World Weather Online*.

Quadro 4 – Descrição das variáveis obtidas por meio da API do *World Weather Online*

Variável	Descrição
temperatura	Temperatura em graus Celsius
umidade	Umidade relativa do ar em %
pressao	Pressão atmosférica em milibars
indice_uv	Índice UV na escala de 0 a 10
sensacao_termica	Sensação térmica em graus Celsius
precipitacao_mm	Quantidade de chuva em mm
velocidade_vento	Velocidade do vento em Km/h
descricao_clima	Descrição da condição climática na data e hora da atividade
dir_vento_graus	Direção do vento em graus
dir_vento	Direção do vento em coordenadas cartesianas

Fonte: Elaborado pelo autor

O Quadro 5 apresenta as variáveis obtidas por meio da API do serviço *LocationIQ*.

Quadro 5 – Descrição das variáveis obtidas por meio da API do *LocationIQ*

Variável	Descrição
cidade	Cidade onde foi realizada cada atividade coletada
estado	Estado ou região à qual pertence a cidade
país	País de origem da atividade coletada

Fonte: Elaborado pelo autor

As variáveis referentes à qualidade de vida foram coletadas conforme o instrumento WHOQOL Abreviado (Anexo C).

O Quadro 6 apresenta as variáveis adicionais, obtidas por meio da anamnese com perguntas complementares ao questionário WHOQOL Abreviado (Apêndice H), além de variáveis calculadas.

Quadro 6 – Descrição das variáveis complementares

Variável	Descrição
data_nascimento	Data de nascimento do participante
idade	Idade do participante em anos, calculada a partir da data de nascimento
sexo	Masculino ou feminino
desempenho_ride	Desempenho (melhorou, piorou, manteve) calculado por meio do cálculo de tendência ao longo do período de prática das atividades (Mann-Kendall)
satisf_sono	O quanto está satisfeito com o sono (satisfeito, insatisfeito, etc.)
satisf_saude	O quanto está satisfeito com a saúde (satisfeito, insatisfeito, etc.)
como_avalia_qv	Como avalia a qualidade de vida (boa, ruim, etc.)
estatura	Estatura em cm
massa_corporal	Massa corporal em kg
preferencia_pratica	Como prefere praticar (individual ou em grupo)
ambiente_pratica	Preferência de ambiente (<i>indoor</i> ou <i>outdoor</i>)
perfil_pratica	Perfil de prática de atividade física (recreativo, amador, profissional)
objetivo	Objetivo com a prática de atividade física (saúde, desempenho, socialização, competição)
som_pref	Pratica ouvindo alguma coisa? (música, podcast, audiolivro, outros, nada)
diabetes	Diagnóstico de diabetes? (sim / não)
hipertensao	Diagnóstico de hipertensão? (sim / não)
cardiaco	Diagnóstico de problema cardíaco? (sim / não)
historico_cardiaco	Histórico cardíaco na família? (sim / não)
uso_medicamento	Usa medicamento? (sim / não)
lesao	Histórico de lesões em decorrência da prática de atividade física? (articulação/tendão/ligamento, muscular, fratura, outras, não)
desconforto	Percepção de desconforto durante a prática de atividade física? (dor nas costas, dor no peito, tontura, outras, não)

Fonte: Elaborado pelo autor

4.4 Variáveis complementares

A combinação de algumas variáveis coletadas possibilita a criação de novas variáveis, por meio de testes estatísticos e equações estabelecidas pela literatura científica. Para o presente estudo, foram adicionadas as seguintes variáveis calculadas: Índice de Massa Corporal (IMC), Índice de Conforto Térmico, scores de qualidade de vida, frequência cardíaca máxima estimada, *pace*³⁰, zonas de intensidade, direção do percurso e desempenho.

O IMC foi calculado de acordo com a equação 2, considerando a massa corporal em quilogramas e a estatura em centímetros.

$$IMC_{kg/m^2} = \frac{\text{massa}}{\text{estatura}^2} \quad (2)$$

O índice de conforto térmico foi calculado conforme a equação da temperatura efetiva em função do vento (equação 3) (PALLOTTA; HERDIES; GONÇALVES, 2015), sendo “ T_a ” a temperatura em graus Celsius, “ UR ” a umidade relativa do ar em percentual e “ v ” a velocidade do vento em metros por segundo.

$$TE_v = 37 - \frac{(37 - T_a)}{0,68 - 0,0014 \times UR + \frac{1}{(1,76 + 1,4 \times v^{0,75})}} - 0,29 \times T_a \times \left(1 - \frac{UR}{100}\right) \quad (3)$$

A compilação do resultado do instrumento WHOQOL Abreviado é feita utilizando o score calculado para cada um dos quatro domínios (físico, psicológico, relações sociais e ambiental), sendo “ $Q_{[x]}$ ” a resposta para cada questão “ x ” identificada de forma numérica nas equações a seguir.

O cálculo do score referente ao domínio físico leva em conta a soma de 7 das 24 facetas, conforme a equação 4.

³⁰ Pace representa o tempo gasto, em minutos, para percorrer uma determinada distância, em quilômetros

$$Df = \frac{\sum Q_{[3,4,10,15,16,17,18]}}{7} \quad (4)$$

Para o domínio psicológico utiliza-se a soma de 6 das 24 facetas, conforme a equação 5.

$$Dp = \frac{\sum Q_{[5,6,7,11,19,26]}}{6} \quad (5)$$

Para o domínio das relações sociais utiliza-se a soma de 3 das 24 facetas, conforme a equação 6.

$$Ds = \frac{\sum Q_{[20,21,22]}}{3} \quad (6)$$

E, por fim, para o domínio ambiental utiliza-se a soma de 8 das 24 facetas, conforme a equação 7.

$$Da = \frac{\sum Q_{[8,9,12,13,14,23,24,25]}}{8} \quad (7)$$

A frequência cardíaca máxima estimada, em batimentos por minuto, foi calculada conforme a equação 8 (TANAKA; MONAHAN; SEALS, 2001)

$$FC_{max} = 208 - 0,7 \times idade \quad (8)$$

O *pace* pode ser calculado conforme a equação 9.

$$P_{min/km} = \frac{Tempo_{min}}{Distância_{km}} \quad (9)$$

As zonas de intensidade foram definidas de acordo com as equações 10, 11, 12 e 13 (CINTIA; PAPPALARDO; PEDRESCHI, 2013) representando,

respectivamente, os intervalos de frequência cardíaca para as zonas de intensidade. As zonas de intensidade são determinadas conforme o Quadro 7.

$$\text{LimiteZ}_1 = FC_{min} + (FC_{max} - FC_{min}) \times 0,6 \quad (10)$$

$$\text{LimiteZ}_2 = FC_{min} + (FC_{max} - FC_{min}) \times 0,7 \quad (11)$$

$$\text{LimiteZ}_3 = FC_{min} + (FC_{max} - FC_{min}) \times 0,8 \quad (12)$$

$$\text{LimiteZ}_4 = FC_{min} + (FC_{max} - FC_{min}) \times 0,9 \quad (13)$$

Quadro 7 – Valores de referência para as zonas de treinamento

Zona de intensidade	Intervalo de frequência cardíaca
Z1	< LimiteZ ₁
Z2	≥ LimiteZ ₁ e < LimiteZ ₂
Z3	≥ LimiteZ ₂ e < LimiteZ ₃
Z4	≥ LimiteZ ₃ e < LimiteZ ₄
Z5	≥ LimiteZ ₄

Fonte: Adaptado de Cintia, Pappalardo e Pedreschi (2013)

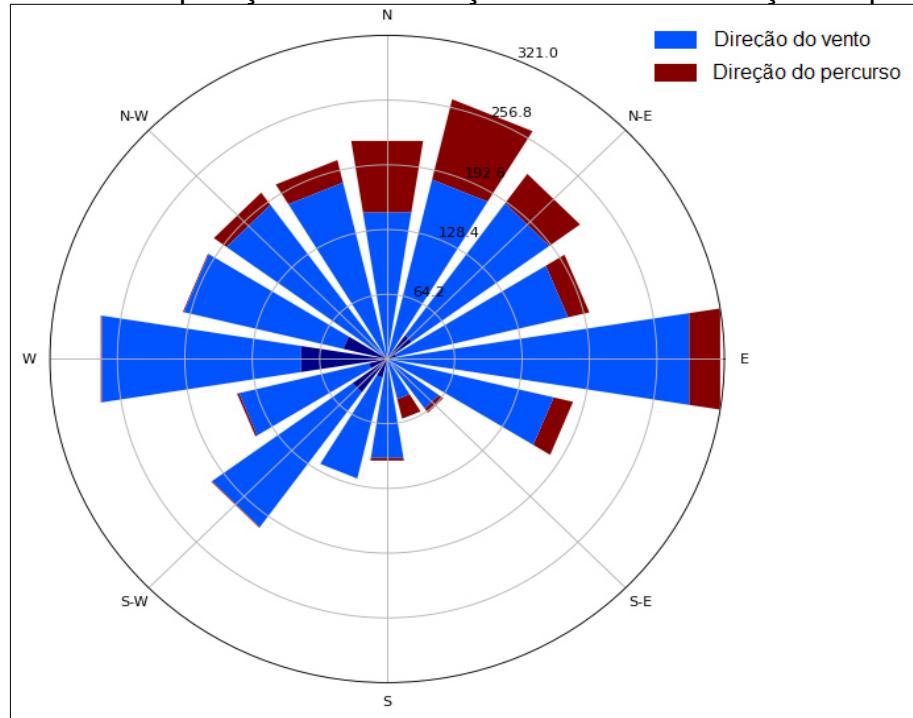
A direção do percurso referente a cada atividade registrada (*bearing*) foi calculada, para cada *trackpoint*³¹, por meio da equação 14, de modo a identificar as variações na direção do percurso, sendo “ φ_1 ” e “ λ_1 ” a latitude e longitude do ponto inicial, “ φ_2 ” e “ λ_2 ” a latitude e longitude do ponto final, “ θ ” a direção em radianos e “ β ” a direção convertida em graus.

$$\begin{aligned}
 y &= \sin(\lambda_2 - \lambda_1) \times \cos(\varphi_2) \\
 x &= \cos(\varphi_1) \times \sin(\varphi_2) - \sin(\varphi_1) \times \cos(\varphi_2) \times \cos(\lambda_2 - \lambda_1) \\
 \theta &= \text{atan} 2(x, y) \\
 \beta &= \theta \times \frac{180}{\pi}
 \end{aligned} \quad (14)$$

31 *Trackpoints* são registros estimados de posicionamento geográfico, baseados em latitude e longitude, determinados geralmente por um equipamento de GPS em um intervalo de tempo regular (CHALKO, 2007).

O bearing é mensurado a partir da direção Norte, em sentido horário (por exemplo, 0° = Norte, 90° = Leste, 180° = Sul e 270° = Oeste). O cálculo da direção do percurso possibilitou a comparação com a direção do vento. A Figura 14 exemplifica tal comparação em um dos percursos analisados, sendo que o maior percentual de coincidência ocorreu na direção leste.

Figura 14 – Comparação entre a direção do vento e a direção do percurso

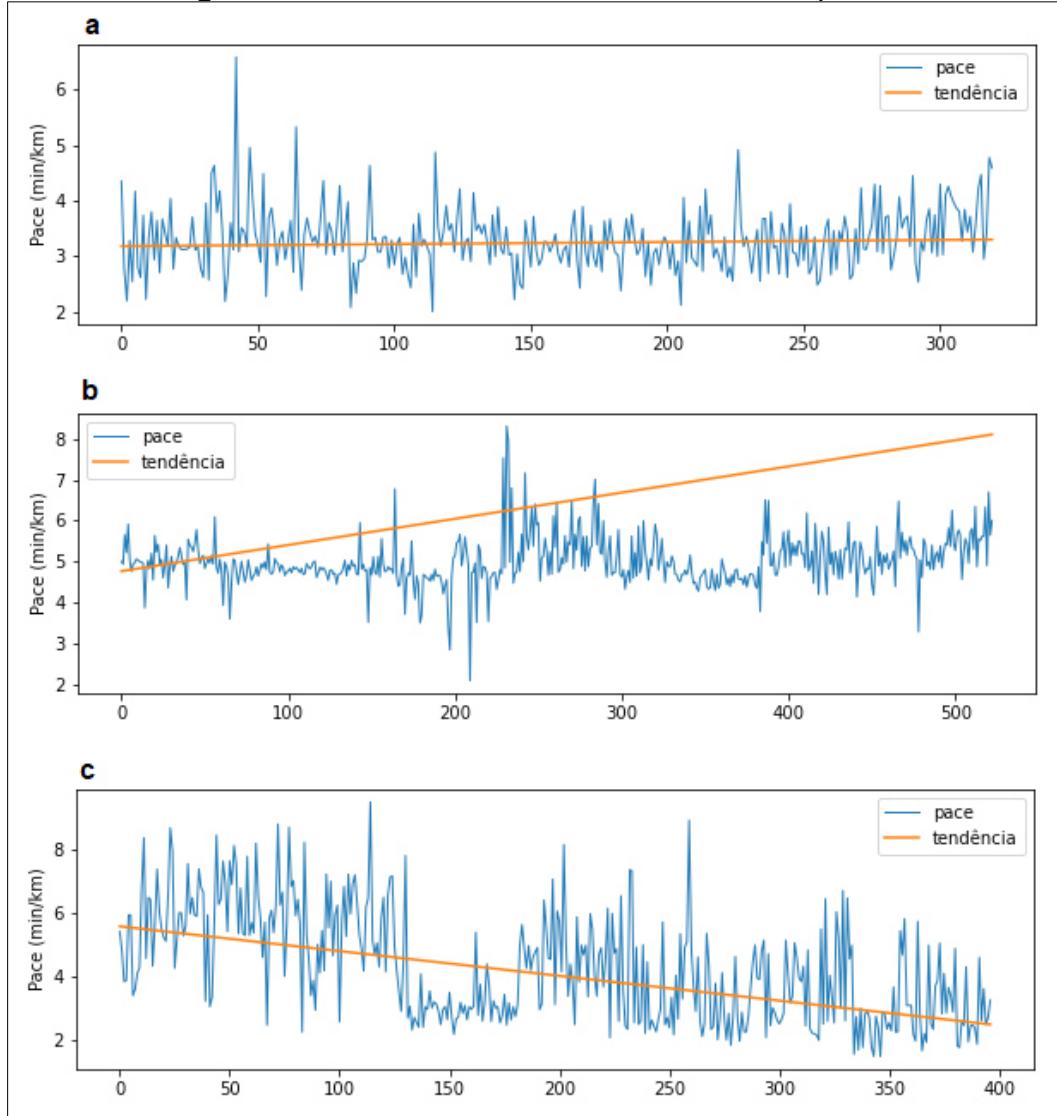


Fonte: Elaborado pelo autor

O desempenho do participante foi avaliado por meio do teste estatístico de tendência de Mann-Kendall para séries temporais, para a variável *pace*, utilizando-se a biblioteca *pyMannKendall*³² e a linguagem de programação *Python*. A Figura 15 exemplifica o comportamento desta variável em uma série temporal, em três situações diferentes: a) sem tendência de alteração no desempenho; b) tendência de redução do desempenho, com o aumento do *pace*; c) tendência de aumento do desempenho, com a redução do *pace*.

32 Disponível em <https://pypi.org/project/pymannkendall>

Figura 15 – Tendência relacionada ao desempenho



Exemplos de gráficos de tendência da variável pace, indicando a manutenção do desempenho:
a) sem tendência

- sem tendência
- redução do desempenho, com tendência de aumento do pace
- aumento do desempenho, com tendência de redução do pace

Fonte: Elaborado pelo autor

4.5 Tabelas de referência

Além das variáveis calculadas, foram utilizadas tabelas padronizadas para classificação do IMC e do conforto térmico.

A classificação do IMC foi realizada conforme o Quadro 8 (NATIONAL INSTITUTES OF HEALTH, 1998).

Quadro 8 – Classificação do IMC

Classificação	IMC (kg/m^2)
Abaixo do peso	< 18,5
Normal	18,5 a 24,9
Sobrepeso	25,0 a 29,9
Obesidade grau I	30,0 a 34,9
Obesidade grau II	35,0 a 39,9
Obesidade grau III	$\geq 40,0$

Fonte: National Institutes of Health (1998)

A classificação do conforto térmico foi realizada conforme o Quadro 9 (PALLOTTA; HERDIES; GONÇALVES, 2015).

Quadro 9 – Classificação das zonas de conforto e desconforto térmico de acordo com a temperatura efetiva em função do vento (TEv) calculada

Sensação térmica	TEv
Muito frio	< 13
Frio	≥ 13 e < 16
Frio moderado	≥ 16 e < 19
Ligeiramente frio	≥ 19 e < 22
Confortável	≥ 22 e < 25
Ligeiramente quente	≥ 25 e < 28
Quente moderado	≥ 28 e < 31
Quente	≥ 31 e < 34
Muito quente	≥ 34

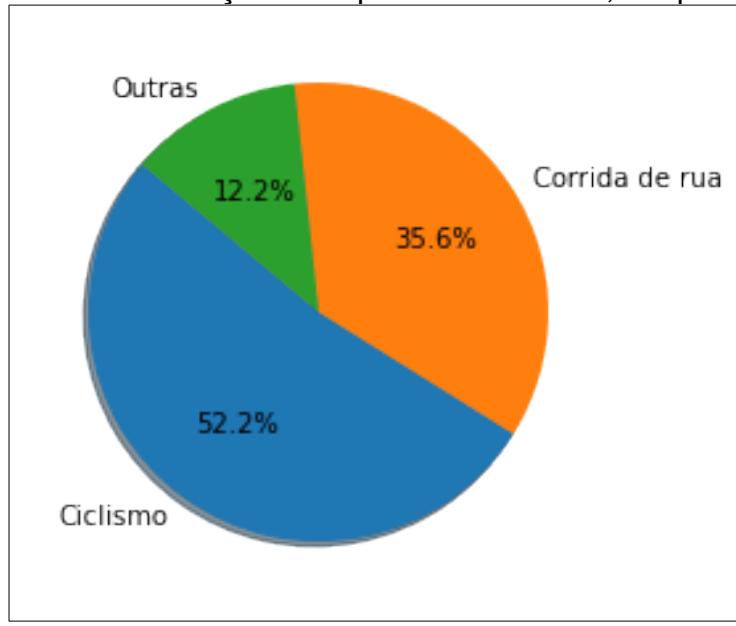
Fonte: Adaptado de Pallotta, Herdies e Gonçalves (2015)

4.6 Limpeza e preparação dos dados

Para o presente estudo foram selecionados somente os dados referentes às atividades de ciclismo e corrida de rua, as quais correspondem à maior parte da

amostra (87,8%), conforme Figura 16. As demais modalidades foram descartadas, pois, cada uma delas representa uma quantidade muito pequena de atividades (apenas 12,2% somadas), o que poderia tornar a análise, usando especificamente a Mineração de Dados, não muito adequada para este estudo.

Figura 16 – Distribuição dos tipos de atividades, em percentual



Fonte: Elaborado pelo autor

De acordo com as diretrizes do ACSM, recomenda-se aos adultos realizarem 30 minutos ou mais de atividade física com intensidade moderada, pelo menos 5 dias por semana (AMERICAN COLLEGE OF SPORTS MEDICINE, 2013). Dessa forma, foram descartadas as atividades com menos de 30 minutos de duração. Quanto à intensidade, não foi necessária nenhuma intervenção nos dados, no sentido de identificar o nível de intensidade de cada atividade realizada, já que tanto o ciclismo quanto a corrida de rua são consideradas atividades de intensidade vigorosa (AMERICAN COLLEGE OF SPORTS MEDICINE, 2013).

O Strava possui um recurso que identifica, de forma automatizada, atividades suspeitas, ou seja, atividades que registraram dados fora dos padrões esperados. Por exemplo, um praticante pode esquecer seu dispositivo ligado e continuar registrando dados após terminar a atividade, com seu veículo em movimento, o que

resultará em dados de velocidade conflitantes com a atividade praticada. Valendo-se deste recurso, 887 atividades marcadas como “suspeitas” foram descartadas.

Por fim, foram removidas também as atividades sem dados de geolocalização e de clima, de modo a restar um conjunto de dados robusto e homogêneo.

4.7 Redução de variáveis

Em aplicações relacionadas à descoberta de conhecimento, de modo geral, uma operação chave é encontrar subconjuntos dos dados que contenham relações suficientes para atender ao foco da análise (ORTEGA, 2000). Ye (2003) cita três diferentes abordagens para a redução de dados e variáveis: criação de novas variáveis a partir do conjunto original; seleção de um subconjunto de variáveis consistindo apenas daquelas que se relacionam; construção de dados a partir do preenchimento de informações ausentes no conjunto original, por exemplo, por meio de inferência. A primeira abordagem foi considerada no presente estudo, por meio do cálculo de variáveis, conforme descrito no item 4.4.

Ao analisar o conjunto original de dados, observou-se a presença de variáveis redundantes e de outras de pouco interesse para esta pesquisa. Assim, também foi aplicado o método para eliminação de variáveis redundantes (RFE) descrito por Kuhn e Johnson (2019). A técnica de RFE se inicia com a criação de um modelo para ser aplicado em todo o conjunto de variáveis, de modo a classificá-las por grau de importância, por meio de um *score* (GUYON et al., 2002). Svetnik et al. (2003) apresentam o algoritmo *Random Forest* como um dos modelos utilizados para tal finalidade.

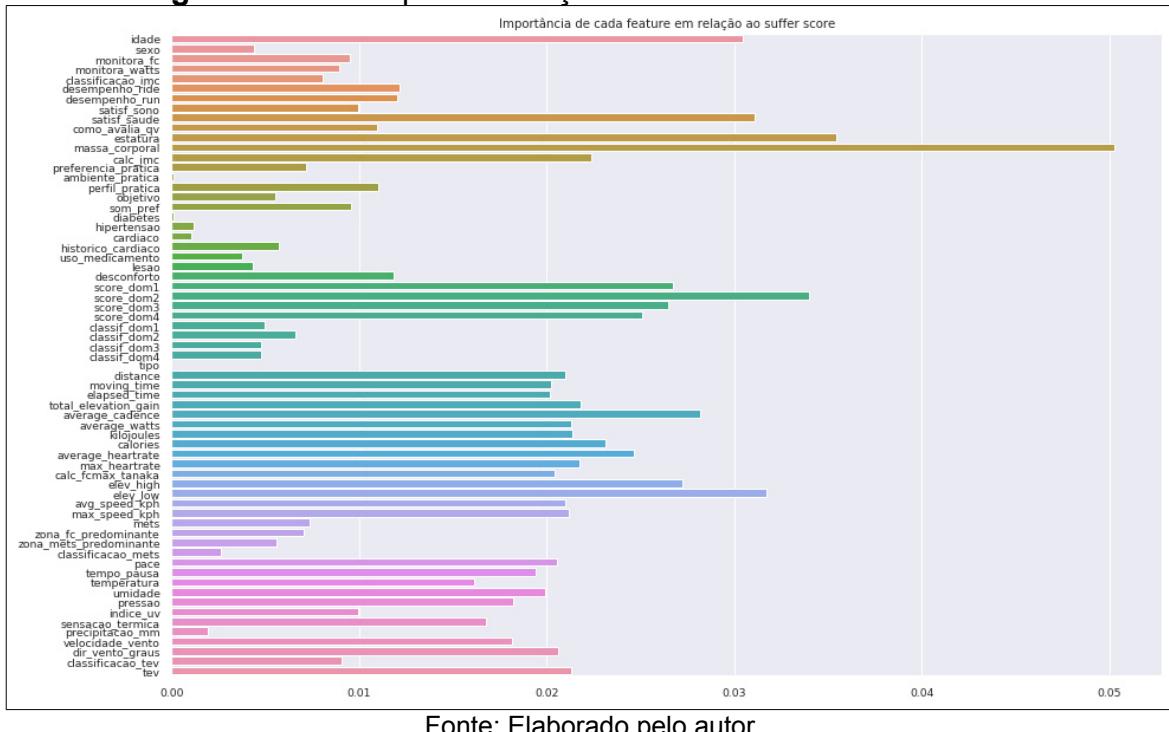
O método empregado no presente estudo consiste em dividir o conjunto de dados em subconjuntos de treinamento e teste e, em seguida, por meio da biblioteca *Scikit-learn*³³ da linguagem de programação *Python*, aplicar o algoritmo *Random Forest*, de modo a obter uma classificação das variáveis por grau de importância em relação a cada variável estudada. A figura 17 apresenta um exemplo de

³³ Disponível em <https://scikit-learn.org>

classificação das variáveis por grau de importância, no caso, aquelas que melhor se relacionam com a variável *suffer score*.

O *suffer score* é uma medida de intensidade disponível no *Strava* que mede a quantidade de esforço necessário para a realização da atividade física, sendo calculada com base na frequência cardíaca (ZENG; ALMQUIST; SPIRO, 2015). No exemplo da Figura 17, por meio da técnica RFE atribui-se um score a cada uma das variáveis. Observa-se no gráfico uma correlação mais elevada para as variáveis “estatura” e “massa corporal”, e uma correlação praticamente inexistente para as variáveis “ambiente de prática” e “tipo de atividade”.

Figura 17 – Exemplo de seleção das variáveis mais relevantes



Fonte: Elaborado pelo autor

4.8 Mineração de Dados

Seguindo o processo de descoberta de conhecimento em bancos de dados descrito por Han, Kamber e Pei (2012), após a limpeza dos dados originais (para remover aqueles inconsistentes e fora dos padrões) e integração dos mesmos (de forma a unificar em um único conjunto os referentes à atividade física, saúde,

qualidade de vida e dados climáticos, resultando em uma base homogênea), passou-se para a etapa de mineração dessa nova base de dados.

Tendo em mãos o conjunto de dados a ser “minerado”, foram estudadas as possibilidades de análise, de forma a permitir a escolha adequada das técnicas e algoritmos a serem empregados; para o desenvolvimento desta pesquisa, a escolhida foi a criação de Árvores de Decisão. Os algoritmos para a criação de árvores de decisão, utilizados no presente estudo, foram implementados em linguagem *Python* e executados no próprio servidor que hospeda o banco de dados. Para fins de verificação do correto funcionamento das implementações realizadas, foram utilizadas também as ferramentas *Weka* versão 3.8 e *R* versão 3.6.2.

Por fim, os padrões encontrados foram analisados visando a obtenção de informações relevantes que auxiliem na tomada de decisão quanto a uma prática mais segura e eficiente da atividade física.

5 ESTUDO EXPERIMENTAL

A fim de mostrar a viabilidade e factibilidade da proposta de pesquisa apresentada, neste capítulo é apresentado um estudo experimental realizado com os dados coletados, contemplando atividades de ciclismo e corrida de rua. Os tópicos deste capítulo abordam a caracterização da amostra, referentes às atividades de ciclismo e corrida de rua, estatísticas e análises realizadas por meio da Mineração de Dados.

5.1 Ciclismo

A prática do ciclismo é bastante popular em diversos países. A bicicleta pode ser utilizada como meio de transporte, para fins recreativos, além de competições nas mais variadas modalidades e diferentes tipos de terreno. Os benefícios para a saúde proporcionados pela prática do ciclismo são amplamente reportados na literatura (DE HARTOG et al., 2010; GÖTSCHI; GARRARD; GILES-CORTI, 2016; OJA et al., 2011), por meio de estudos que apresentam a relação de tal prática com a melhora da aptidão física, a diminuição de doenças cardiovasculares, redução da obesidade, entre outros benefícios. No entanto, vários fatores podem influenciar, tanto de forma positiva quanto de forma negativa, na prática do ciclismo, de modo que devemos observar, por exemplo, limites individuais, condições climáticas e ajustes do equipamento. O presente estudo possibilitou observar várias relações entre variáveis quanto ao desempenho, condições climáticas, percepção de saúde e percepção da qualidade de vida.

5.1.1 Amostra

A amostra utilizada para as análises referentes ao ciclismo é composta por 168.409 atividades registradas no aplicativo *Strava*. A Tabela 2 resume as características da amostra.

Tabela 2 – Caracterização das atividades realizadas (ciclismo)

Distância média percorrida	N	Tempo (minutos)	Distância (km)
Até 10 km	27.466	17 ± 7	6,1 ± 2,5
10 a 40 km	74.309	62 ± 23	24,3 ± 8,5
Acima de 40 km	66.634	169 ± 94	70,9 ± 44,8

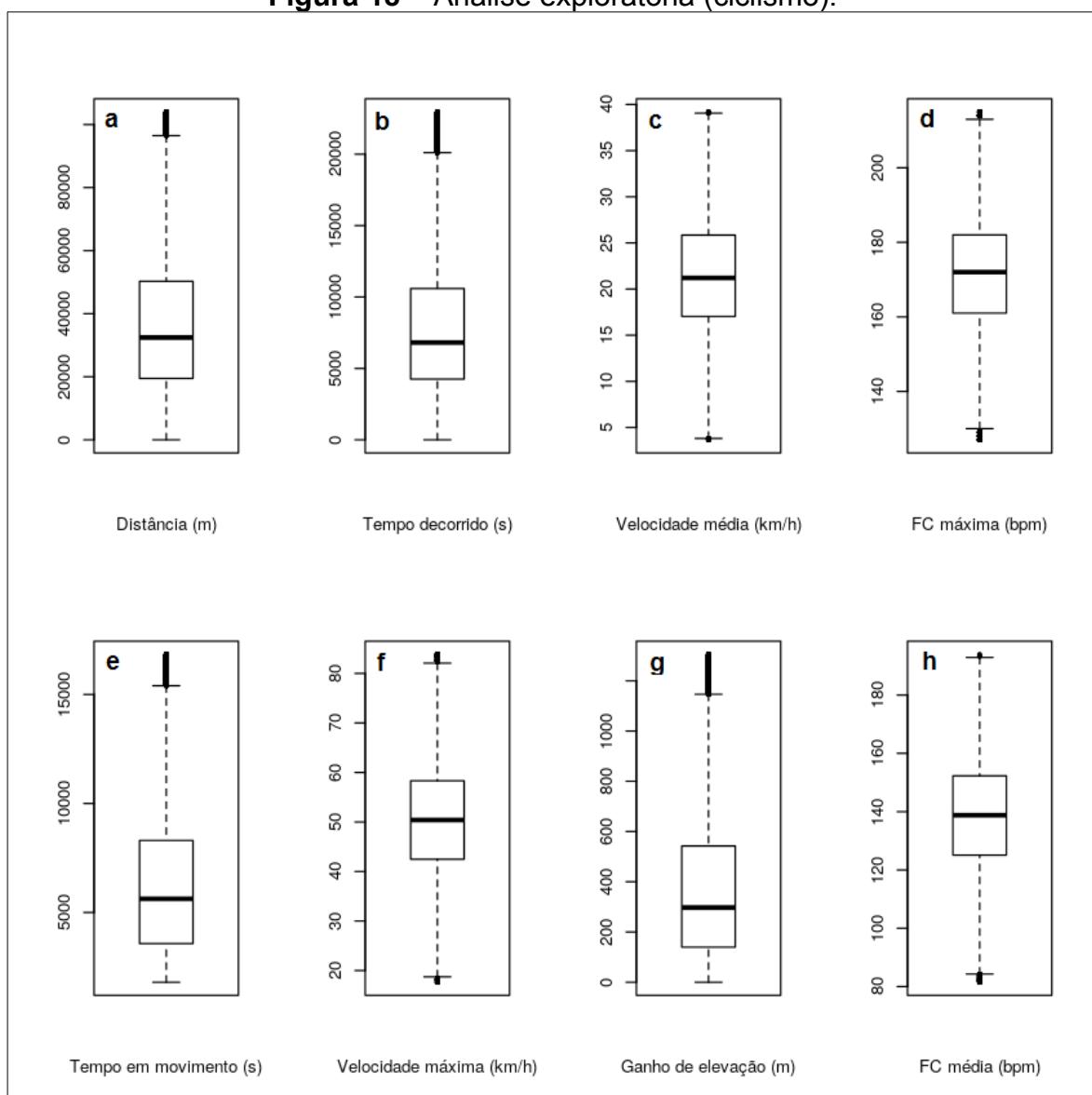
Fonte: Elaborado pelo autor

5.1.2 Análise exploratória e estatísticas

Por meio de uma análise exploratória utilizando gráficos do tipo *boxplot* (Figura 18) observou-se uma presença significativa de *outliers*³⁴. Isto se explica devido a uma grande variação no perfil das atividades, como, por exemplo, distâncias variando de muito curtas a muito longas e variação de velocidade, devido a diferentes perfis de atletas.

³⁴ Em estatística, o termo “*outliers*” diz respeito a valores extremos ou atípicos, observações que apresentam um grande afastamento das demais observações do conjunto de dados estudado (GLADWELL, 2013).

Figura 18 – Análise exploratória (ciclismo).



Gráficos tipo *boxplot* referentes a:

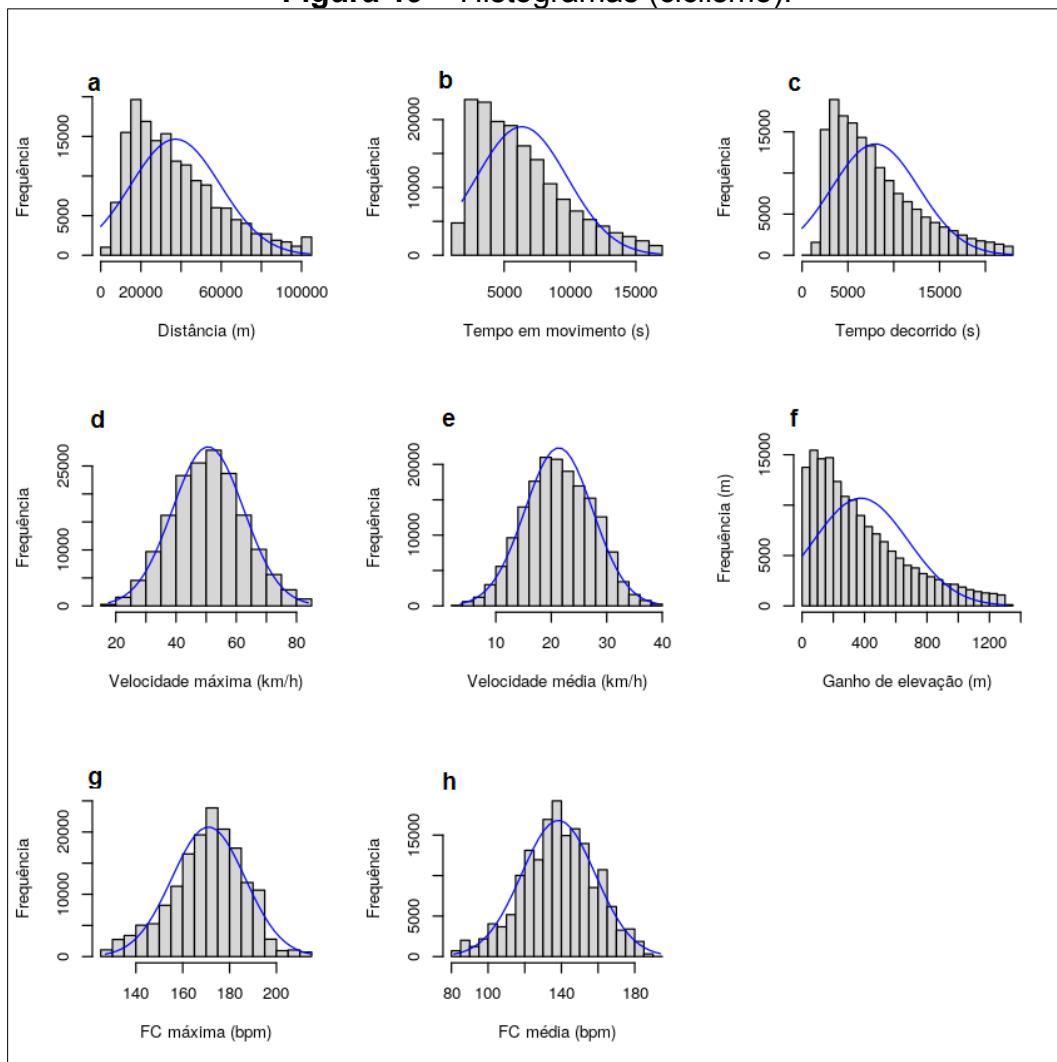
- a) distância
- b) tempo decorrido
- c) velocidade média
- d) frequência cardíaca máxima
- e) tempo em movimento
- f) velocidade máxima
- g) ganho de elevação
- h) frequência cardíaca média

Fonte: Elaborado pelo autor

Tomando como exemplo o gráfico referente ao ganho de elevação (Figura 18-g), tal ganho varia de 0 a 1306 metros, com média de 376 metros e mediana de 297 metros. Observa-se a presença de *outliers*, referentes a casos esporádicos em que o ganho de elevação foi acima de 1100 metros.

Foram realizados testes estatísticos para verificar a normalidade das distribuições. Em um primeiro momento, estes testes foram realizados de forma visual, por meio de gráficos do tipo histograma (Figura 19), e, na sequência, por meio do teste estatístico Shapiro-Wilk, utilizando-se a biblioteca *SciPy*³⁵ e a linguagem de programação *Python*.

³⁵ Disponível em <https://www.scipy.org>

Figura 19 – Histogramas (ciclismo).

Histogramas referentes a:

- distância
- tempo em movimento
- tempo decorrido
- velocidade máxima
- velocidade média
- ganho de elevação
- frequência cardíaca máxima
- frequência cardíaca média

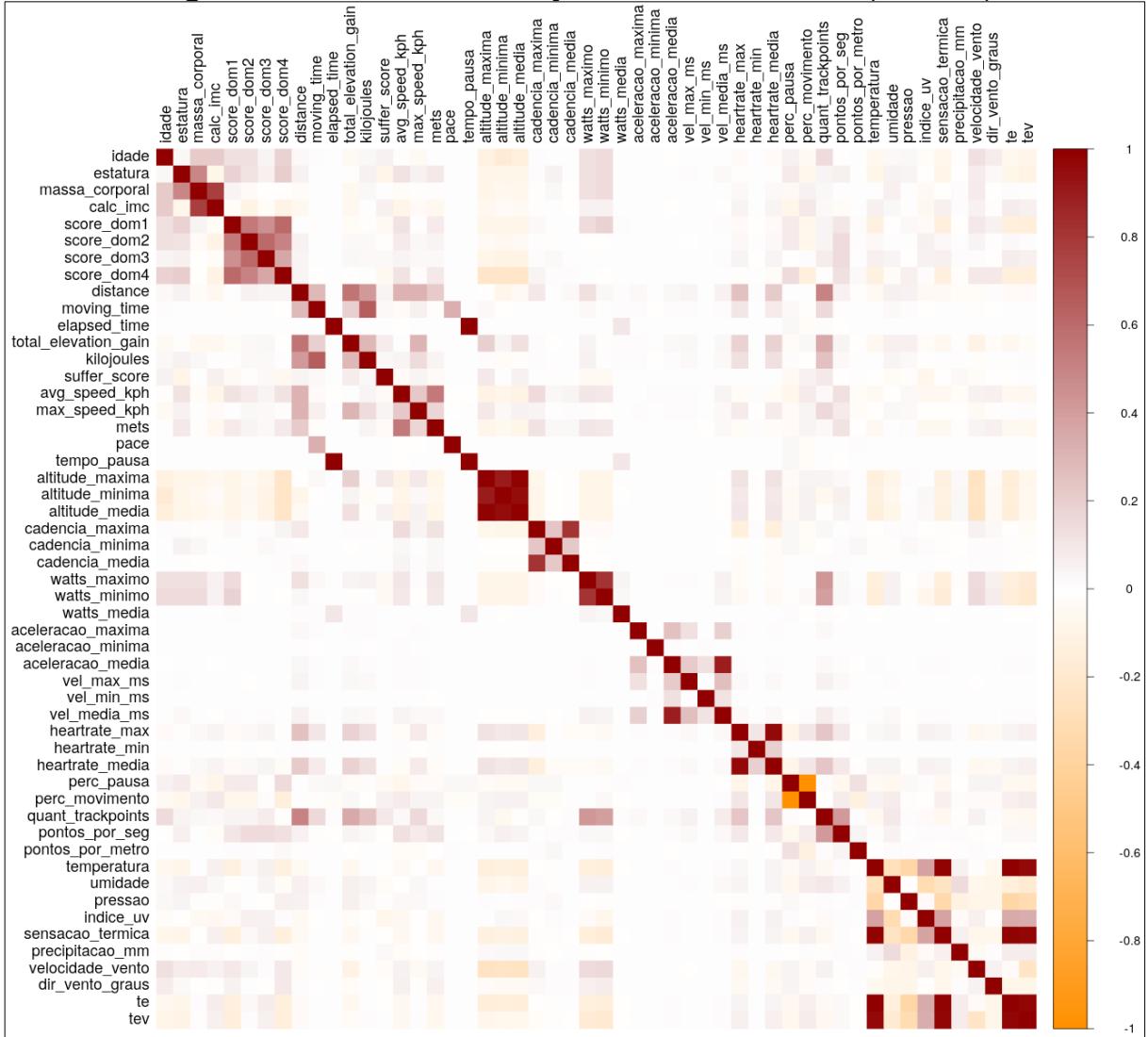
Fonte: Elaborado pelo autor

A normalidade de uma amostra se caracteriza pela distribuição dos valores em torno do valor central (média), de forma homogênea (desvio padrão baixo). Devido à característica da amostra utilizada no presente estudo, composta por atletas de diferentes perfis (faixa etária, condição física, sexo) e atividades com características muito variadas (distância, tempo de duração, tipo de terreno,

condições climáticas), a distribuição dos valores ocorre de forma mais dispersa em relação à média em todas as variáveis observadas, apresentando desvio padrão mais elevado, o que levou à rejeição da hipótese de normalidade para todas as variáveis observadas.

Finalizando a análise exploratória, em uma outra análise visual, agora considerando as correlações entre variáveis (Figura 20), observa-se claramente que estas acontecem somente entre as variáveis de mesmo grupo, ou seja, variáveis climáticas se correlacionam, e o mesmo ocorre com variáveis de qualidade de vida e variáveis das atividades.

Figura 20 – Matriz de correlações entre as variáveis (ciclismo)



Fonte: Elaborado pelo autor

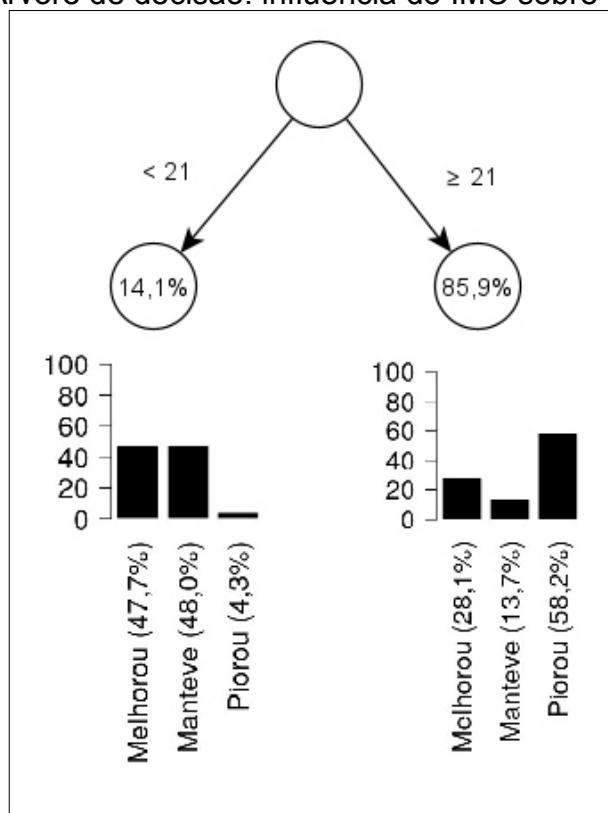
Desse modo, justifica-se a busca de uma abordagem diferente, como a Mineração de Dados, a fim de buscar novas relações entre as variáveis analisadas e obtenção de novos conhecimentos a partir do conjunto de dados coletados.

5.1.3 Mineração de Dados aplicada ao ciclismo

Analizando a influência do Índice de Massa Corporal (IMC) sobre o desempenho de ciclistas (Figura 21), observa-se melhora no desempenho de forma mais significativa (47,7%) quando o IMC ficou abaixo de 21 kg/m^2 . Observa-se, também, a manutenção do desempenho de forma significativa (48%) quando o IMC

ficou abaixo de 21 kg/m^2 . Com este resultado, podemos refletir sobre a necessidade de maior atenção aos valores de IMC quando se busca manter ou melhorar o desempenho, ou seja, não apenas seguir a classificação padrão considerada normal ($18,5$ a $24,9 \text{ kg/m}^2$) (NATIONAL INSTITUTES OF HEALTH, 1998), pois mesmo dentro da faixa considerada normal, a melhora de desempenho se mostrou mais significativa quando os valores estão mais próximos ao limite mínimo de tal faixa.

Figura 21 – Árvore de decisão: influência do IMC sobre o desempenho

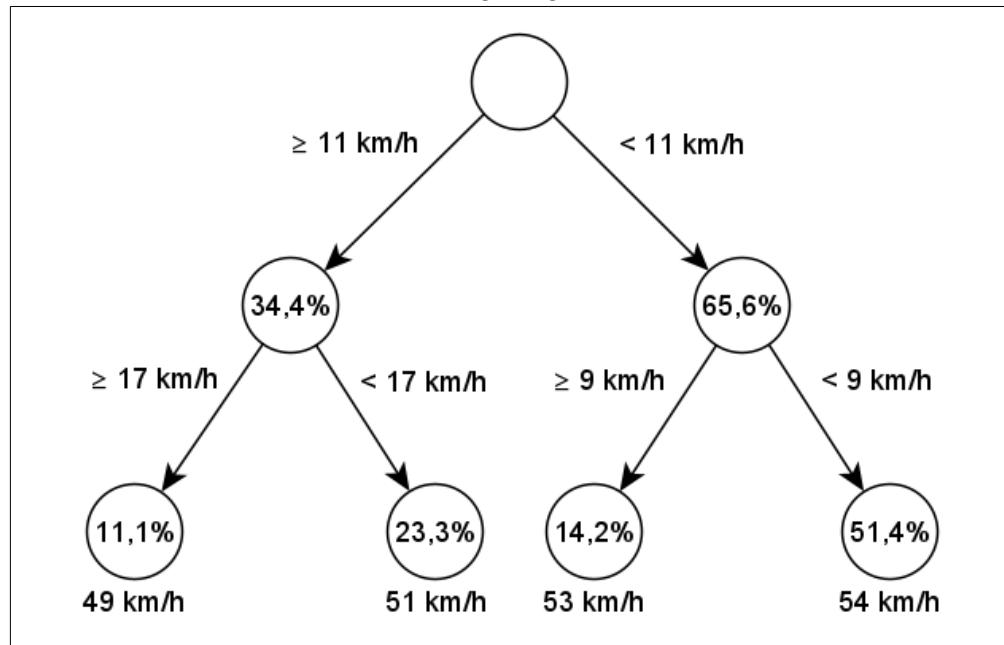


Fonte: Elaborado pelo autor

Outro aspecto relevante observado no presente estudo diz respeito a algumas variáveis climáticas. Analisando a influência da velocidade do vento sobre a velocidade máxima no ciclismo (Figura 22), observa-se que ventos fortes podem afetar significativamente a velocidade máxima obtida. Com ventos acima de 17 km/h , a velocidade máxima alcançada foi de 49 km/h . Por outro lado, para velocidade do vento abaixo de 9 km/h , a velocidade máxima foi de 54 km/h . Dessa forma, podemos inferir que um cenário ideal quando se deseja realizar a atividade com menos esforço seria ventos abaixo de 9 km/h . No entanto, caso o objetivo de

um determinado treino seja realizar maior esforço, ventos com velocidade mais elevada podem ser uma vantagem interessante para atingir o objetivo almejado.

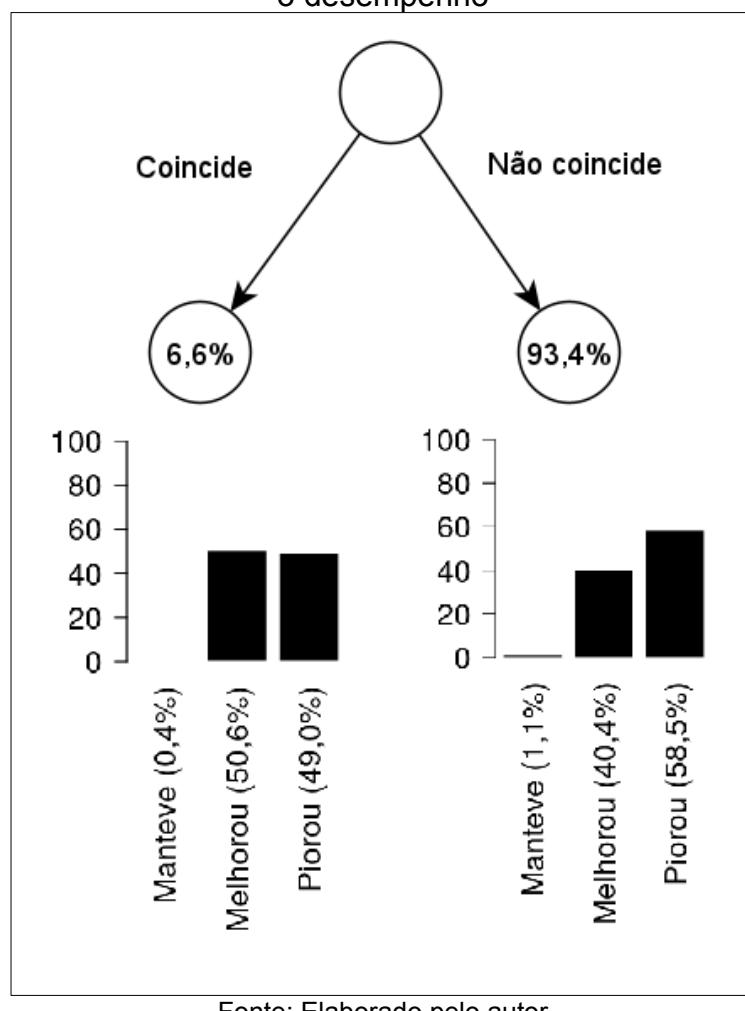
Figura 22 – Árvore de decisão: influência da velocidade do vento sobre a velocidade máxima



Fonte: Elaborado pelo autor

Assim como a velocidade do vento, a direção deste pode impactar na prática do ciclismo (Figura 23). Quando a direção do vento não coincide com a direção do percurso, ou seja, quando temos ventos contrários ou laterais, observa-se uma piora mais acentuada no desempenho (58,5% contra 49% quando a direção do vento coincide). Observa-se também uma melhora mais acentuada no desempenho quando a direção do vento coincide com o percurso (50,6% contra 40,4% quando a direção não coincide).

Figura 23 – Árvore de decisão: influência das direções do vento e do percurso sobre o desempenho

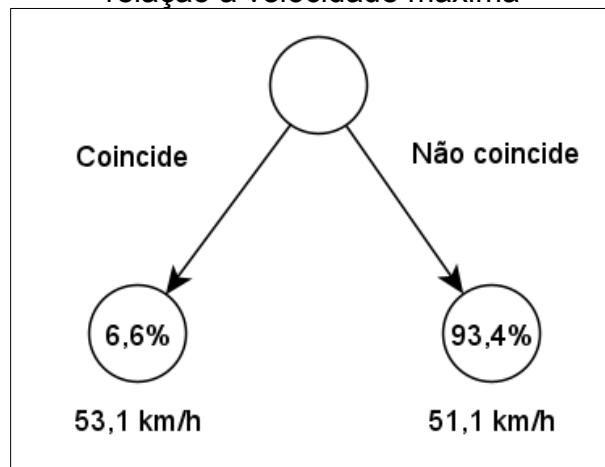


Fonte: Elaborado pelo autor

Analizando a direção do vento em relação à direção do percurso (Figura 24), observa-se que, quando as direções coincidem, o ciclista consegue alcançar uma maior velocidade máxima (53,1 km/h contra 51,1 km/h para ventos contrários).

Embora seja um fenômeno conhecido, a técnica aqui empregada é diferenciada por ser capaz de quantificar tal fenômeno.

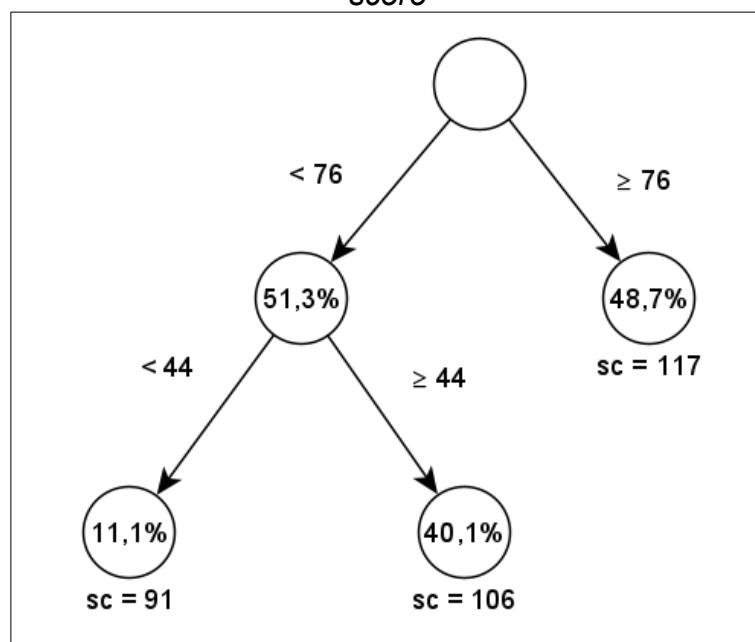
Figura 24 – Árvore de decisão: influência das direções do vento e do percurso em relação à velocidade máxima



Fonte: Elaborado pelo autor

Ainda analisando características climáticas, no caso da influência da umidade relativa do ar sobre o *suffer score* (sc) de ciclistas (Figura 25), observa-se que, quando aferida acima de 76%, contribui para o aumento deste indicador (117), enquanto valores abaixo de 44% sinalizam para uma redução (91).

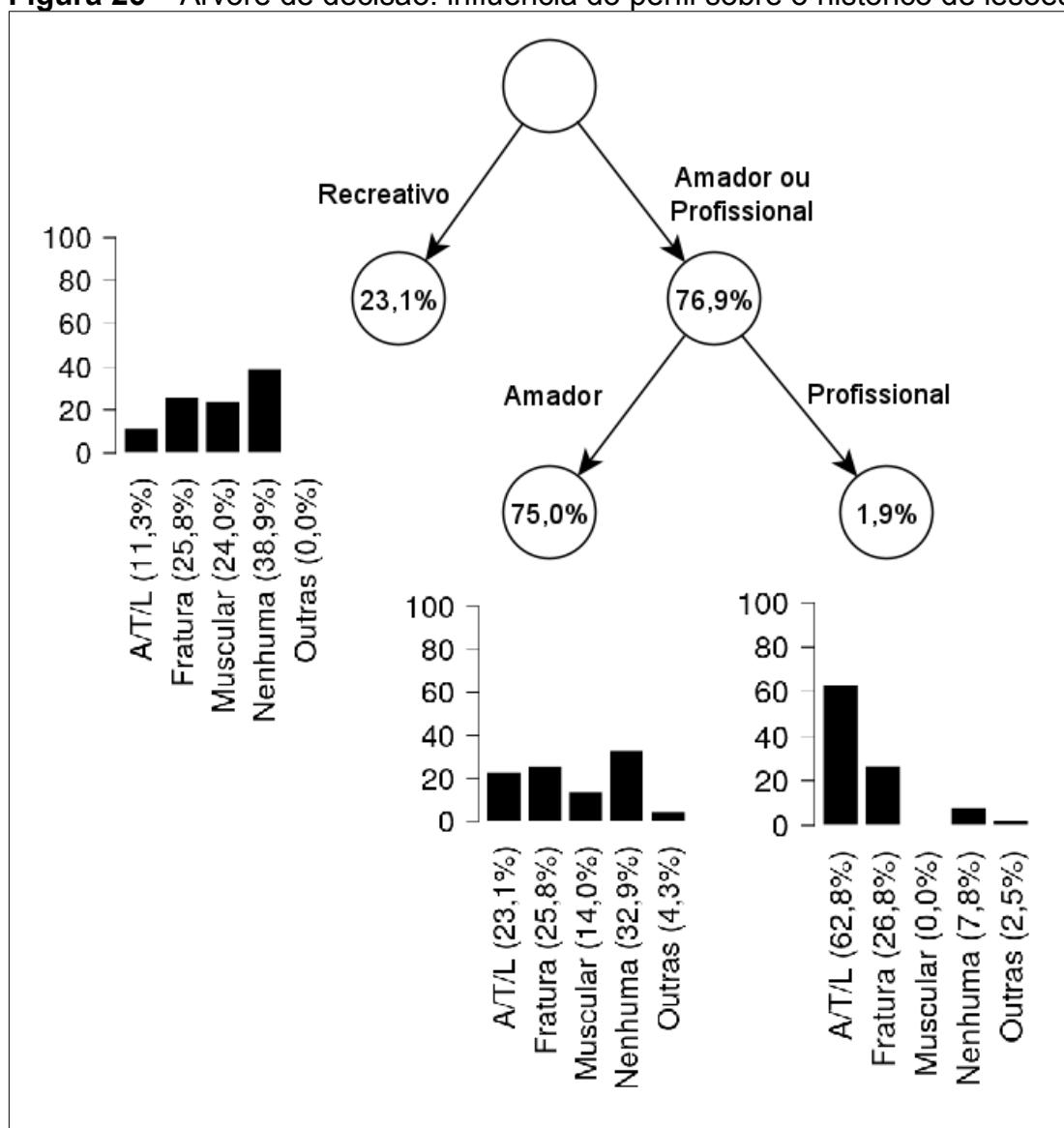
Figura 25 – Árvore de decisão: influência da umidade relativa do ar sobre o *suffer score*



Fonte: Elaborado pelo autor

Na sequência, foi analisada a influência do perfil do praticante de ciclismo sobre o histórico de lesões (Figura 26). Observa-se que as lesões de articulação, tendão ou ligamento (A/T/L) são as de maior ocorrência em praticantes profissionais, enquanto que, no grupo com perfil de prática amador ou recreativo, a incidência de lesões desta natureza é significativamente menor. Dos praticantes pertencentes ao grupo profissional 62,8% relatam já ter sofrido alguma lesão relacionada à articulação, tendão ou ligamento.

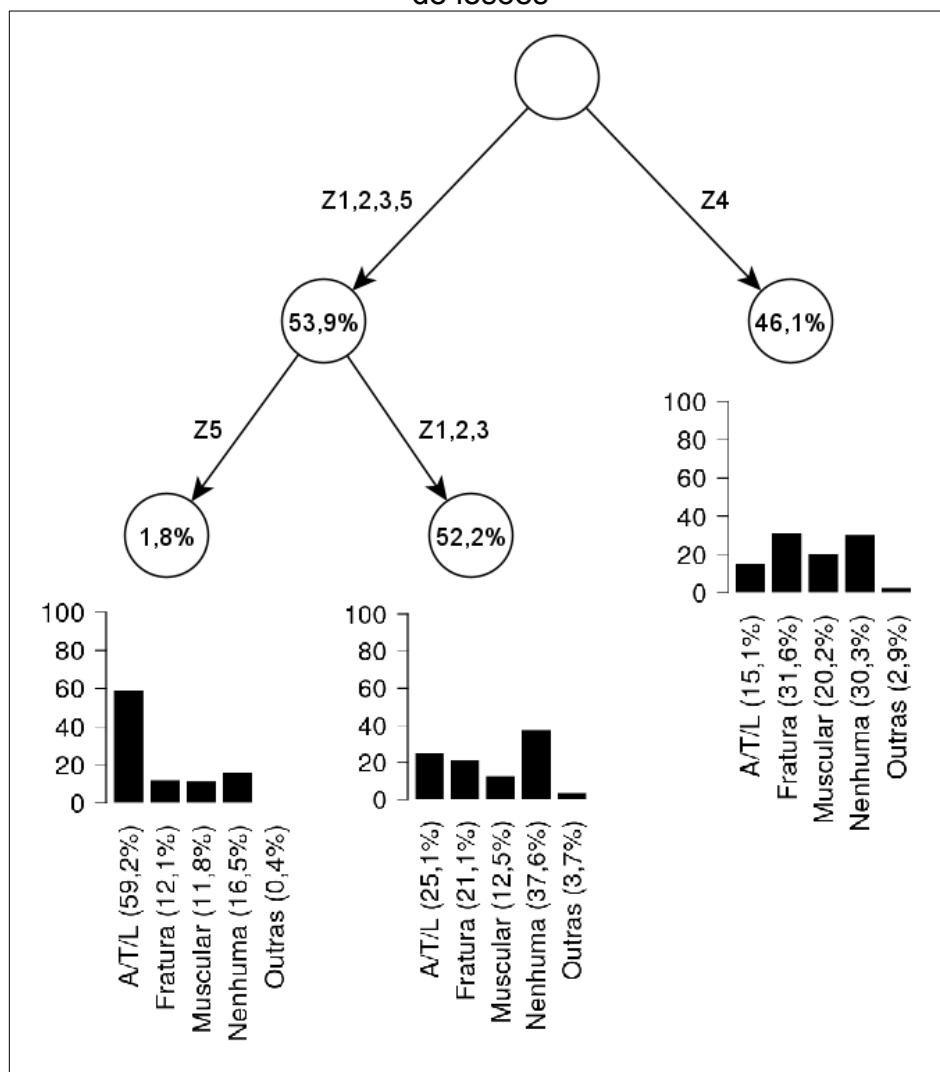
Figura 26 – Árvore de decisão: influência do perfil sobre o histórico de lesões



Fonte: Elaborado pelo autor

Prosseguindo, foi analisada a influência das zonas de intensidade sobre o histórico de lesões em ciclistas (Figura 27). Observa-se que a maior incidência de lesões de articulação, tendão ou ligamento (A/T/L), representando 59,2% dos casos de lesão, está associada ao nível mais elevado de intensidade (Z5). Na zona de intensidade Z5 observa-se que apenas 16,5% não relatam histórico de lesões, enquanto na zona Z4 este percentual aumenta para 30,3%. Já nas zonas de menor intensidade (Z1, Z2 e Z3), o percentual dos que não reportam a ocorrência de lesões cresce ainda mais (37,6%). Estes números reforçariam a existência da relação entre alta intensidade e o histórico de lesões.

Figura 27 – Árvore de decisão: influência das zonas de intensidade sobre o histórico de lesões



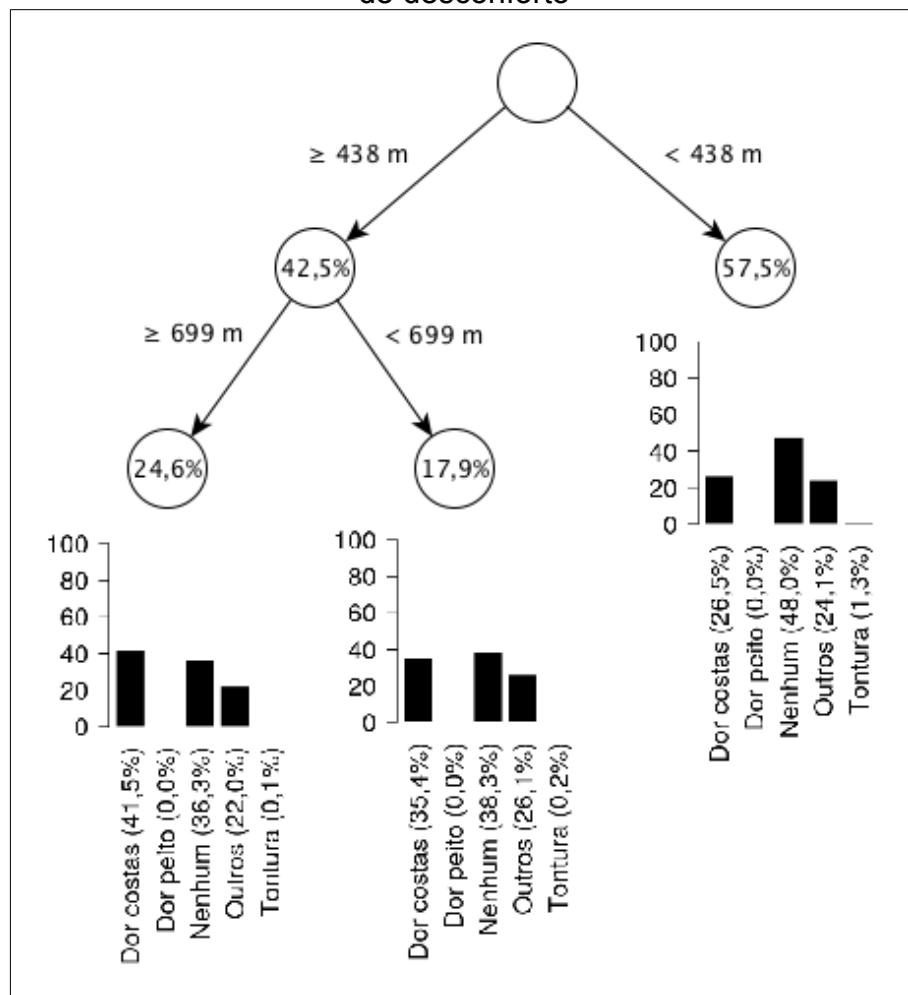
Fonte: Elaborado pelo autor

Nestas zonas de menor intensidade (Z1, Z2 e Z3), somando as lesões do tipo A/T/L (25,1%) com lesões do tipo muscular (12,5%), as lesões não traumáticas representam 37,6%. No caso da zona de intensidade Z4, a maioria (31,6%) reporta lesões do tipo fratura, que são lesões traumáticas. No entanto, somando as lesões do tipo A/T/L (15,1%) com lesões do tipo muscular (20,2%), as lesões não traumáticas representam 35,3% do total. Considerando tanto as lesões traumáticas quanto as não traumáticas, a zona de intensidade Z5 apresenta 83,1% de lesões, a zona de intensidade Z4 apresenta 66,9% de lesões e as zonas de intensidade mais baixa (Z1, Z2 e Z3) apresentam 58,7% de lesões, o que, de novo, reforçaria a relação de alta intensidade com o histórico de lesões.

O joelho é uma das estruturas com maior incidência de lesões não traumáticas no ciclismo (DETTORI; NORVELL, 2006). A Figura 27 mostra que as lesões não traumáticas (A/T/L e lesão muscular) são relatadas por 71% dos praticantes da zona Z5, 35,3% da zona Z4 e por 37,6% das zonas de menor intensidade. Baseado nisso, e corroborando com Ruiz e Oliveira (2008), deve-se refletir sobre a importância de um treinamento visando o fortalecimento muscular, de forma a proteger as articulações (além de uma atenção especial quanto ao ajuste dos equipamentos, tais como bicicleta e sapatilhas).

Os dados gerados quanto à influência do ganho de elevação sobre a percepção de desconforto de ciclistas destacam dois valores importantes para tal ganho: 438 metros e 699 metros. A Figura 28 mostra que quando o ganho de elevação fica abaixo de 438 metros, 26,5% dos participantes relataram dor nas costas. Quando este ganho de elevação ficou entre 438 metros e 699 metros, 35,4% relatam dor nas costas. Prosseguindo a análise, quando o ganho de elevação fica acima de 699 metros, o percentual dos que relatam dor nas costas sobe para 41,5%. O que se observa é que a incidência de dor nas costas aumenta proporcionalmente com aumento do ganho de elevação.

Figura 28 – Árvore de decisão: influência do ganho de elevação sobre a percepção de desconforto

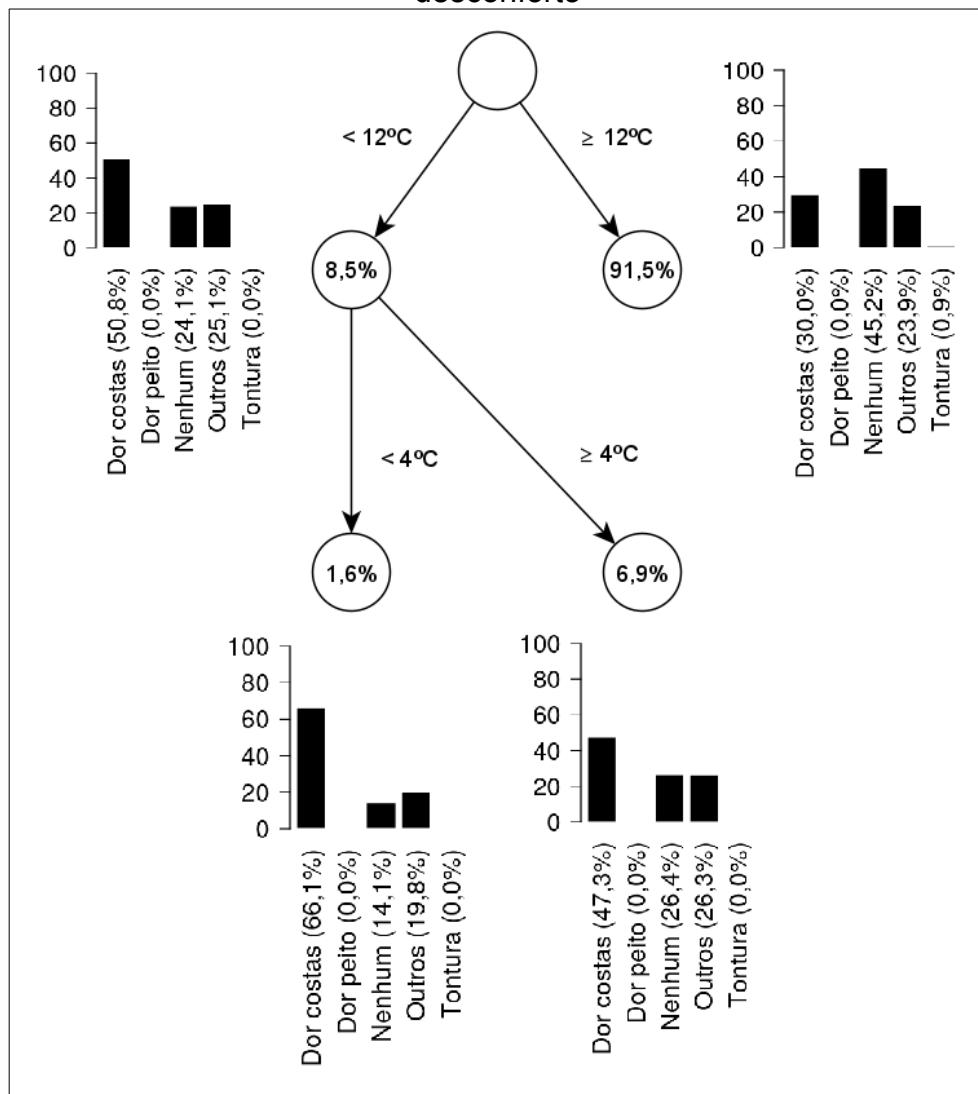


Fonte: Elaborado pelo autor

Dor no pescoço e dor nas costas são comuns em ciclistas devido à posição do corpo durante a atividade (ASPLUND; WEBB; BARKDULL, 2005), principalmente durante as subidas quando, geralmente, os ciclistas alternam entre posição sentado e em pé (DUC et al., 2008). Períodos prolongados de subidas e fadiga muscular contribuem para a dor nas costas em ciclistas (SCHULZ, 2009). Os resultados apresentados acima evidenciam a relação entre o acúmulo de subidas e a incidência de dor nas costas. Nesse caso, o aumento no ganho de elevação deve ser gradual e concomitante com um trabalho de fortalecimento muscular da região lombar, de modo que o condicionamento físico do praticante seja compatível com o acúmulo de subidas durante cada atividade.

Ainda analisando a percepção de desconforto em ciclistas, observa-se que a temperatura (Figura 29), quando abaixo dos 12º C, também é fator relevante, principalmente no que diz respeito a dor nas costas. Em temperaturas abaixo de 12º C, a maioria (75,9%) relata algum tipo de desconforto, sobressaindo a incidência de dor nas costas (50,8%), enquanto 24,1% não reportam nenhum tipo de desconforto. Em temperaturas abaixo de 4º C, a incidência de algum tipo de desconforto aumenta para 85,9%, destacando-se a dor nas costas (66,1%). Em temperaturas acima de 12º C, a incidência de desconforto cai para 54,8%, predominando dor nas costas com 30%, e o percentual que não relata nenhum desconforto aumenta para 45,2%.

Figura 29 – Árvore de decisão: influência da temperatura sobre a percepção de desconforto

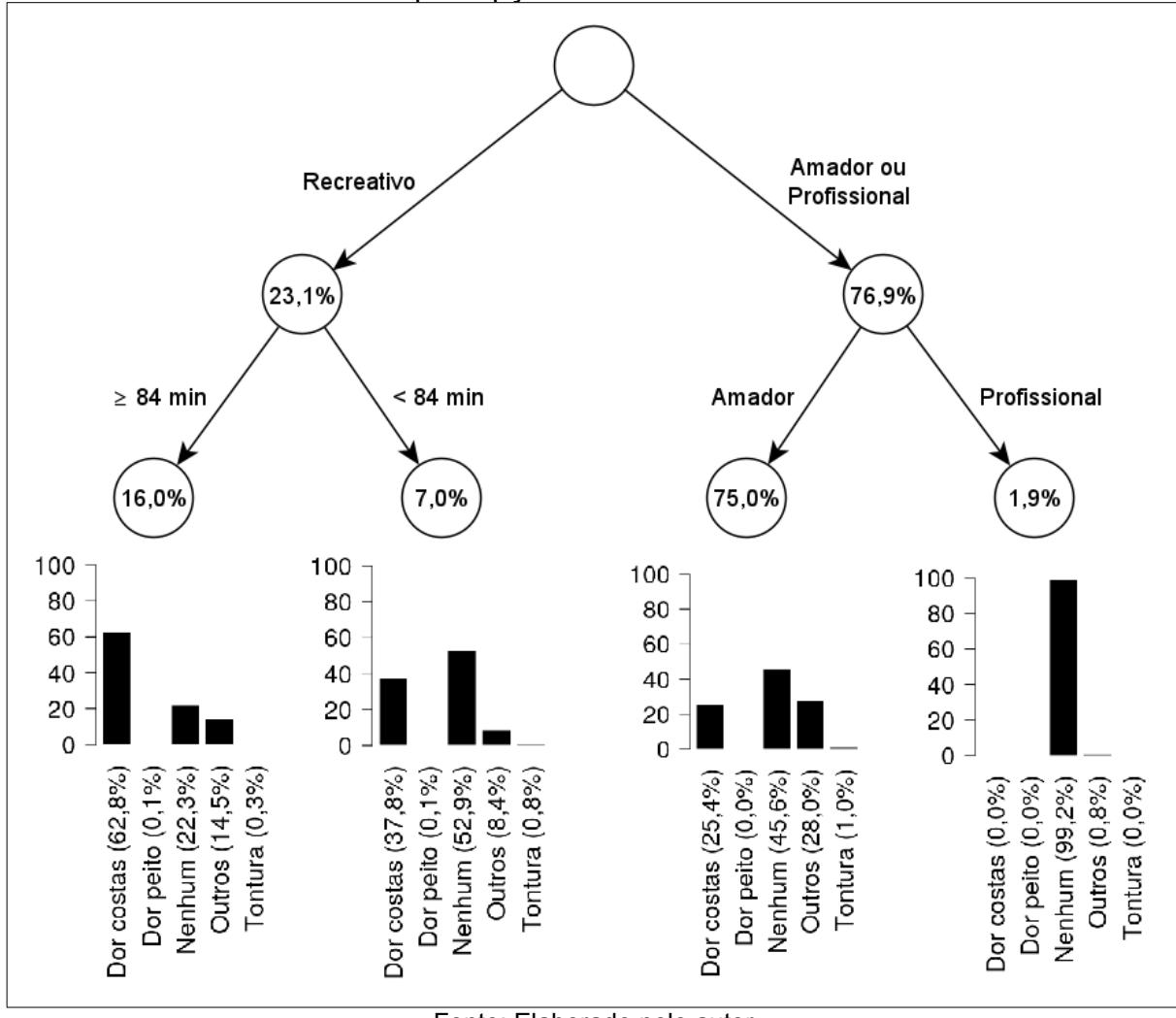


Fonte: Elaborado pelo autor

A prática do ciclismo pode ocorrer nos mais diversos ambientes e localizações, com as mais variadas condições climáticas, como frio, calor, variações de altitude, vento, chuva, entre outras. Principalmente praticantes inexperientes podem sofrer com uma rápida fadiga em decorrência de mudanças climáticas (ANSARI; NOURIAN; KHODAEE, 2017). No caso de baixas temperaturas, o consumo energético aumenta na tentativa de manter o corpo aquecido, o que também pode levar a uma condição de fadiga ou mesmo exaustão. Portanto, levando-se em consideração as possíveis condições climáticas, recomenda-se o uso de equipamentos e vestuário adequados, além de um planejamento prévio da atividade.

Outro aspecto que pode ser observado nos resultados, quando o foco é voltado para o perfil do praticante de ciclismo, é quanto à influência do tempo total decorrido, incluindo tempo de movimento e pausas, sobre a percepção de desconforto (Figura 30), com destaque para a incidência de dor nas costas. Quando o perfil de prática é amador ou profissional, a maioria não relata nenhum tipo de desconforto (45,6% dos amadores e 99,2% dos profissionais), enquanto, dos amadores, 25,4% reportam dor nas costas e 28% outros tipos de desconforto. A incidência de desconforto é insignificante quando se trata do perfil profissional. No caso dos praticantes recreativos, o tempo decorrido parece influenciar de forma significativa na percepção de desconforto. Para estes, quando o tempo decorrido fica igual ou superior a 84 minutos, a maioria (62,8%) relata dor nas costas, 22,3% não percebe nenhum tipo de desconforto e 14,5% reportam outros tipos de desconforto. Quando o tempo decorrido fica abaixo de 84 minutos, uma porcentagem bem menor dos praticantes recreativos reportam dor nas costas (37,8%), enquanto 52,9% não percebem nenhum tipo de desconforto. Independentemente do tempo decorrido, o praticante recreativo sempre apresentou uma maior incidência de dor nas costas quando comparado ao amador ou profissional (62,8% quando o tempo é maior que 84 minutos e 37,8% quando o tempo é inferior a isso, em contraste com 25,4% do praticante amador e 0,0% do atleta profissional), provavelmente devido a um pior condicionamento da musculatura na região lombar.

Figura 30 – Árvore de decisão: influência do perfil e do tempo decorrido sobre a percepção de desconforto



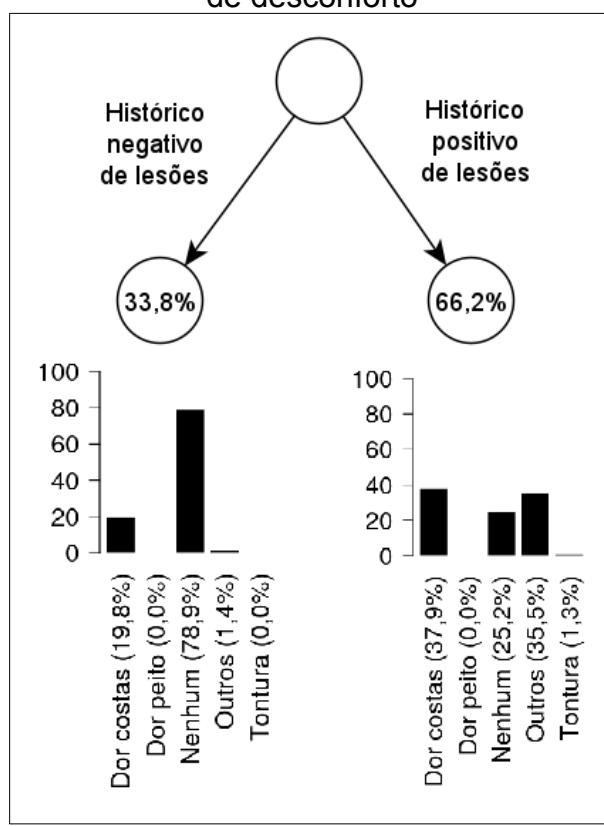
Fonte: Elaborado pelo autor

Atletas amadores e profissionais, de modo geral, possuem um condicionamento físico melhor quando comparados a praticantes recreativos, pois tendem a encarar a prática da atividade de forma mais séria e disciplinada. A posição do ciclista na bicicleta leva a uma flexão prolongada das costas, o que pode resultar em dor muscular quando a estrutura está mal condicionada. A região lombar é composta pelo principal grupo muscular que gera energia e controla o movimento da bicicleta. Se as costas não estiverem bem condicionadas e flexíveis, ocorrem fadiga e tensão muscular, causando dor (ASPLUND; WEBB; BARKDULL, 2005). Dessa forma, o tempo de duração da atividade deve ser proporcional ao nível de

condicionamento físico do praticante, destacando-se a importância de um trabalho de fortalecimento muscular, principalmente na região lombar.

A Figura 31 apresenta a influência do histórico de lesões sobre a percepção de dor nas costas. Quando o praticante não reporta nenhum tipo de lesão (histórico negativo), a maioria (78,9%) também não reporta nenhum tipo de desconforto, enquanto 19,8% reportam dor nas costas. Quando o praticante relata histórico de lesão (histórico positivo), a maioria (37,9%) relata também a ocorrência de dor nas costas, enquanto 25,2% não relatam nenhum tipo de desconforto e 35,5% relatam desconforto de outros tipos.

Figura 31 – Árvore de decisão: influência do histórico de lesões sobre a percepção de desconforto



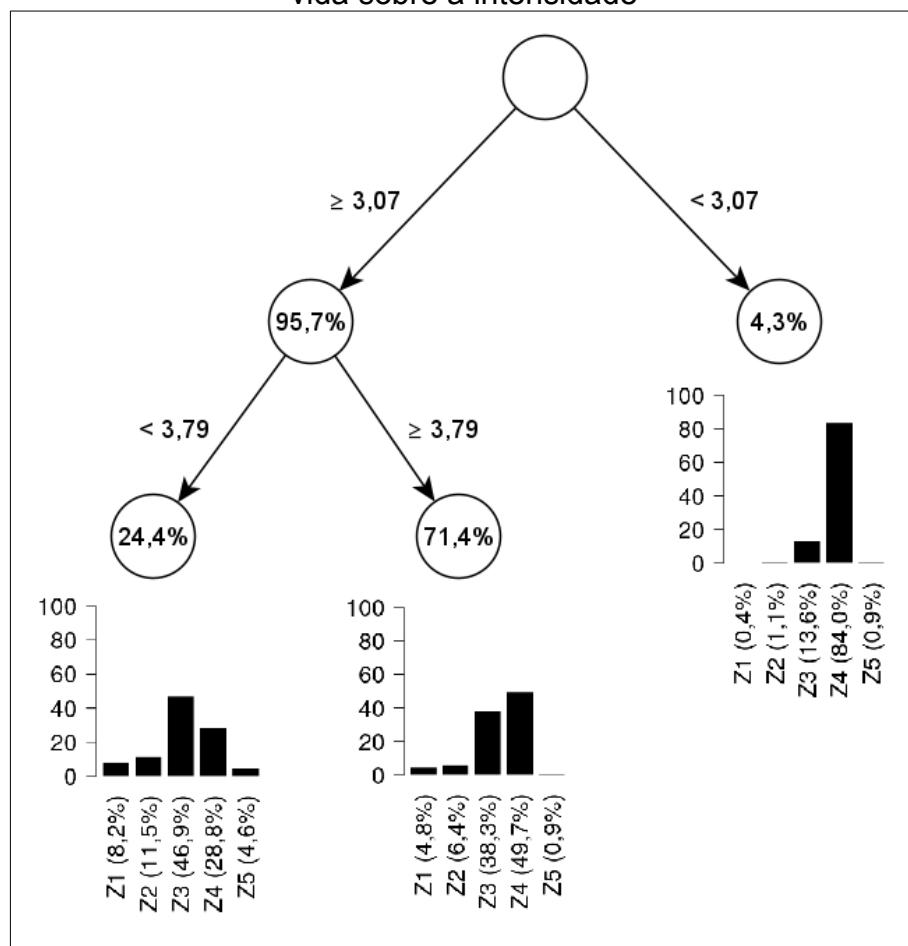
Fonte: Elaborado pelo autor

Não foram encontradas na literatura evidências quanto aos motivos da relação entre desconforto e histórico de lesões no ciclismo. No entanto, podemos refletir sobre o fator psicológico envolvido: devido ao receio de que a lesão se repita, o praticante pode, involuntariamente, adotar uma postura não natural, levando a tal

desconforto. Ainda, pode ser decorrente de fator físico, no caso de uma lesão mal curada que pode levar a uma postura adaptada, na tentativa de compensar um desconforto decorrente da lesão prévia, levando a outro tipo de desconforto (WEINBERG; GOULD, 2017).

Outro ponto observado dentro deste estudo foi a influência do score do domínio físico do questionário WHOQOL Abreviado sobre a intensidade (Figura 32). O que se observa é que, quando o score fica abaixo de 3,07, as atividades são realizadas quase sempre na zona 4 de intensidade (84%), o que demonstraria um desequilíbrio quanto às bases científicas do treinamento esportivo, principalmente no que diz respeito a volume e intensidade (DANTAS, 2014). Por outro lado, quando o score fica acima de 3,07, observa-se um treinamento realizado de forma um pouco mais equilibrada, com predominância nas zonas de intensidade Z3 e Z4.

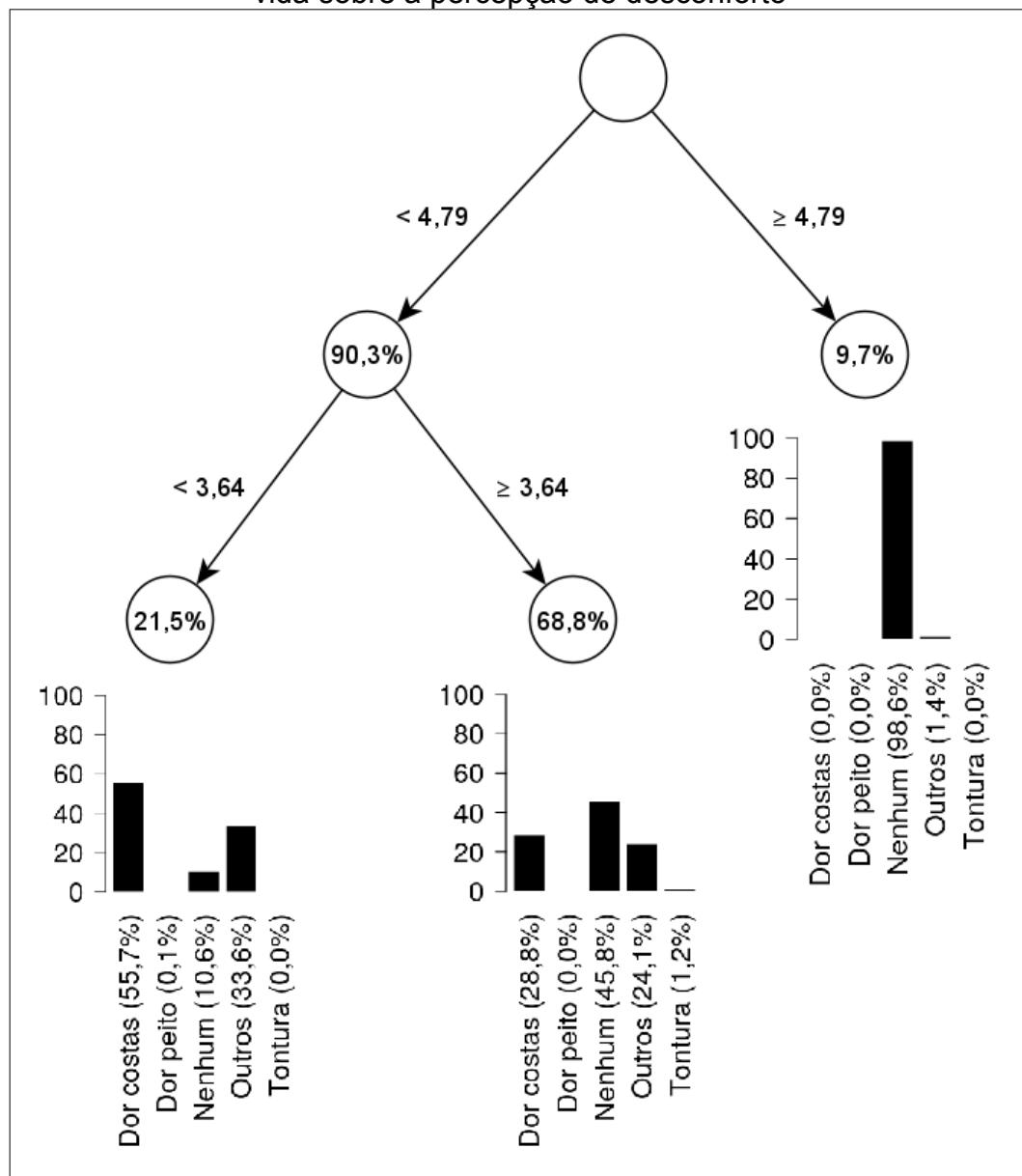
Figura 32 – Árvore de decisão: influência do score do domínio físico de qualidade de vida sobre a intensidade



Fonte: Elaborado pelo autor

Com a análise do score de ciclistas, referente ao domínio físico (Figura 33) avaliado pelo questionário WHOQOL Abreviado, observa-se o aumento do percentual dos que não reportam nenhum desconforto proporcionalmente ao aumento do score. Quando o score está abaixo de 3,64, apenas 10,6% não reportam nenhum tipo de desconforto, enquanto a soma dos demais representa 89,4%, com destaque para dor nas costas, com 55,7%. Quando o score fica acima de 4,79, esse percentual aumenta para 98,6%, de modo que a percepção de desconforto, de qualquer natureza, representa apenas 1,4%.

Figura 33 – Árvore de decisão: influência do score do domínio físico de qualidade de vida sobre a percepção de desconforto



Fonte: Elaborado pelo autor

5.2 Corrida de rua

A corrida de rua é uma das práticas esportivas mais populares em todo o mundo, contando com mais de 10 milhões de praticantes no Brasil, o que pode ser explicado pela facilidade de aprendizado e baixo custo (HINO et al., 2009). Assim como ocorre no ciclismo, os benefícios para a saúde proporcionados pela prática da corrida de rua também são amplamente reportados na literatura, com estudos (LAVIE et al., 2015; LEE et al., 2017) apresentando sua relação com a melhora da aptidão física, a diminuição de doenças cardiovasculares, redução da obesidade, entre diversos outros benefícios (MORGAN; COSTILL, 1996). Como acontece em qualquer outra modalidade esportiva, diversos fatores podem impactar, tanto de forma positiva quanto de forma negativa, a prática da corrida de rua, de modo que devemos observar as mais diversas variáveis envolvidas nesta atividade. O presente estudo possibilitou observar várias relações entre variáveis relacionadas a desempenho, condições climáticas, percepção de saúde e percepção da qualidade de vida.

5.2.1 Amostra

A amostra utilizada para as análises referentes à corrida de rua é composta por 72.009 atividades registradas no aplicativo *Strava*. A Tabela 3 resume as características da amostra.

Tabela 3 – Caracterização das atividades realizadas (corrida de rua)

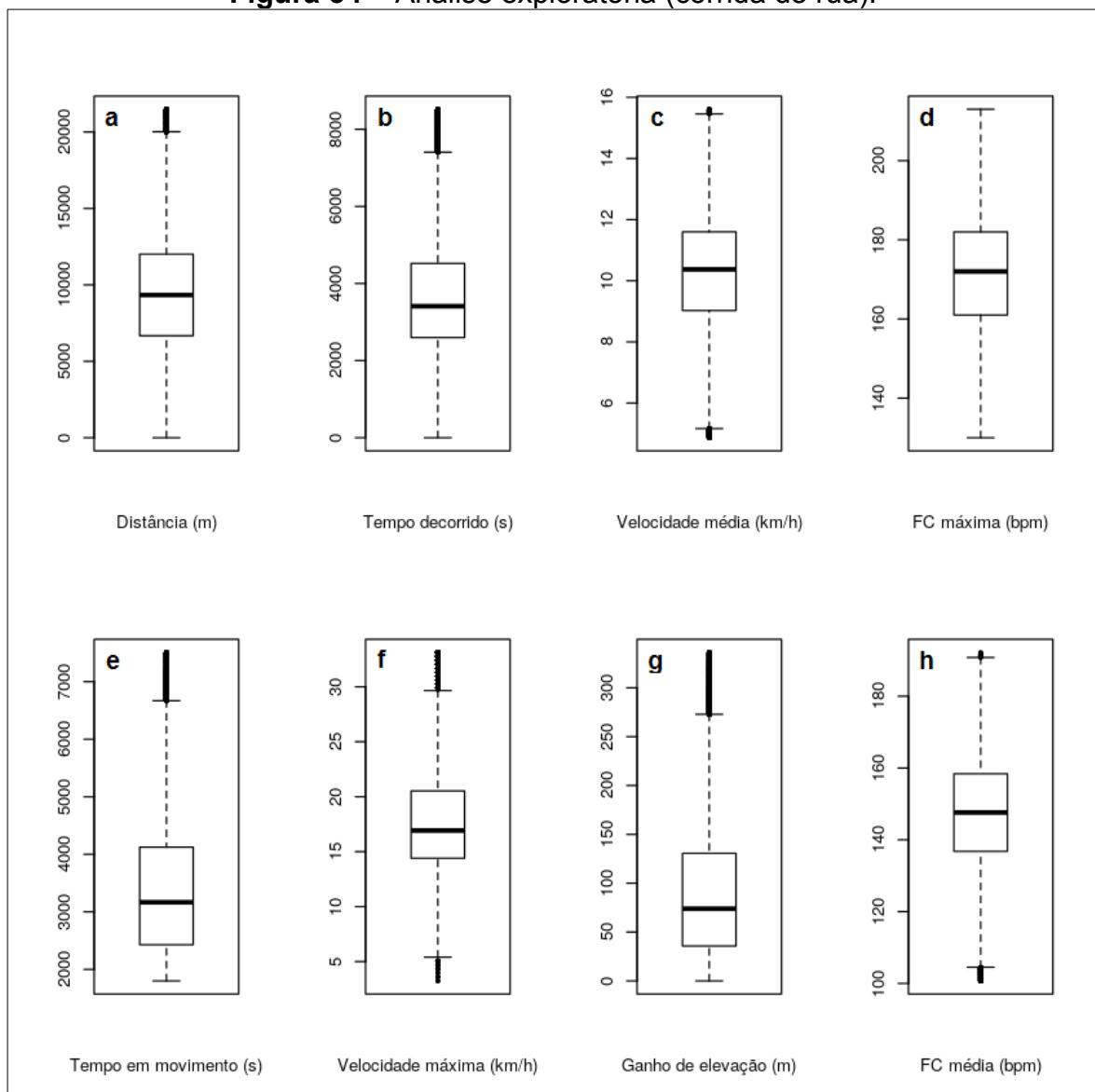
Distância média percorrida	N	Tempo (minutos)	Distância (km)
Até 5 km	15.500	12 ± 8	2,6 ± 1,5
5 a 20 km	52.149	49 ± 18	9,9 ± 3,5
Acima de 20 km	4.360	124 ± 33	25,3 ± 6,2

Fonte: Elaborado pelo autor

5.2.2 Análise exploratória e estatísticas

Assim como no ciclismo, a análise exploratória (Figura 34) indica uma presença significativa de *outliers*.

Figura 34 – Análise exploratória (corrida de rua).



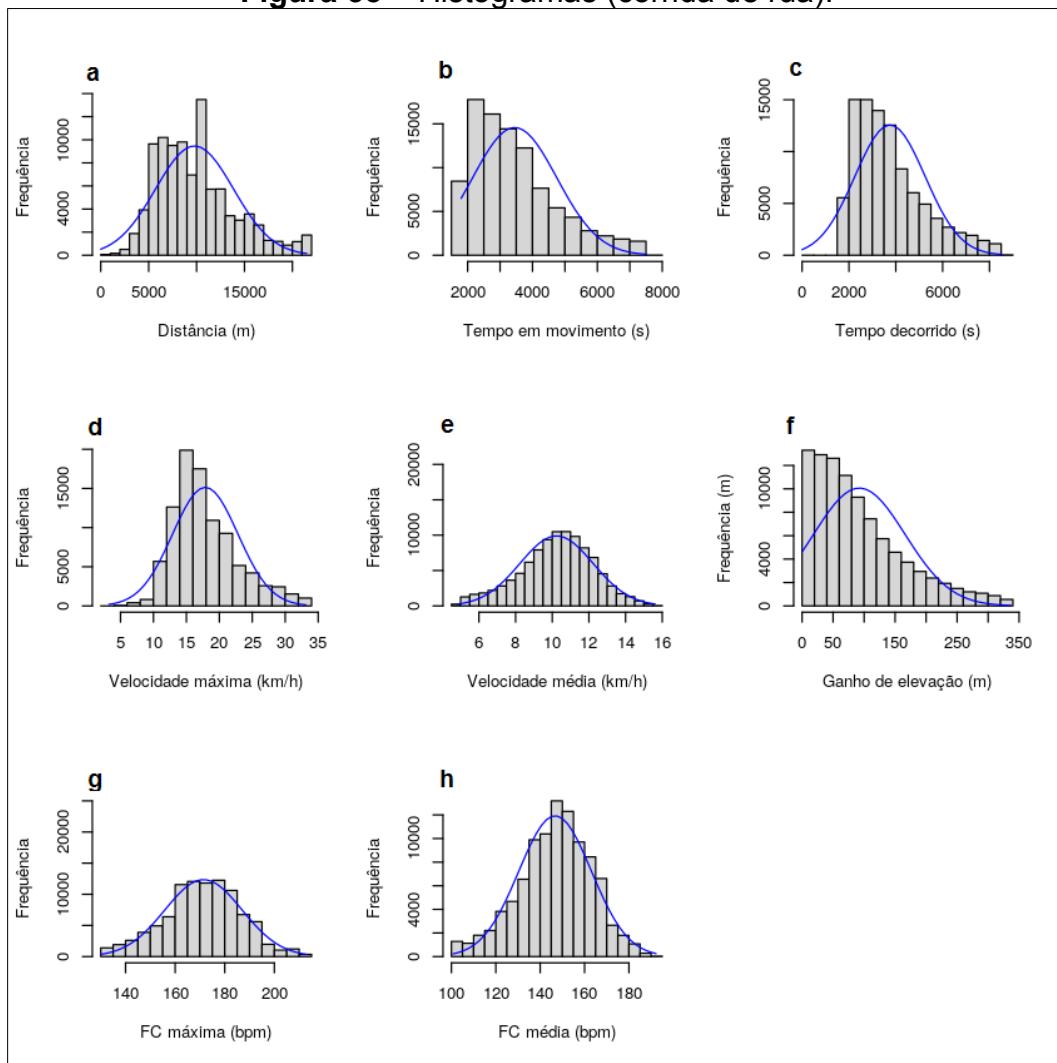
Gráficos tipo *boxplot* referentes a:

- distância
- tempo decorrido
- velocidade média
- frequência cardíaca máxima
- tempo em movimento
- velocidade máxima
- ganho de elevação
- frequência cardíaca média

Fonte: Elaborado pelo autor

Também aqui foram realizados testes estatísticos para verificar a normalidade das distribuições. Em um primeiro momento, estes testes foram realizados de forma visual, por meio de histogramas (Figura 35), e, na sequência, por meio do teste estatístico Shapiro-Wilk, utilizando-se a biblioteca *SciPy* e a linguagem de programação *Python*. Assim como no ciclismo, a hipótese de normalidade foi rejeitada para todas as variáveis observadas.

Figura 35 – Histogramas (corrida de rua).



Histogramas referentes a:

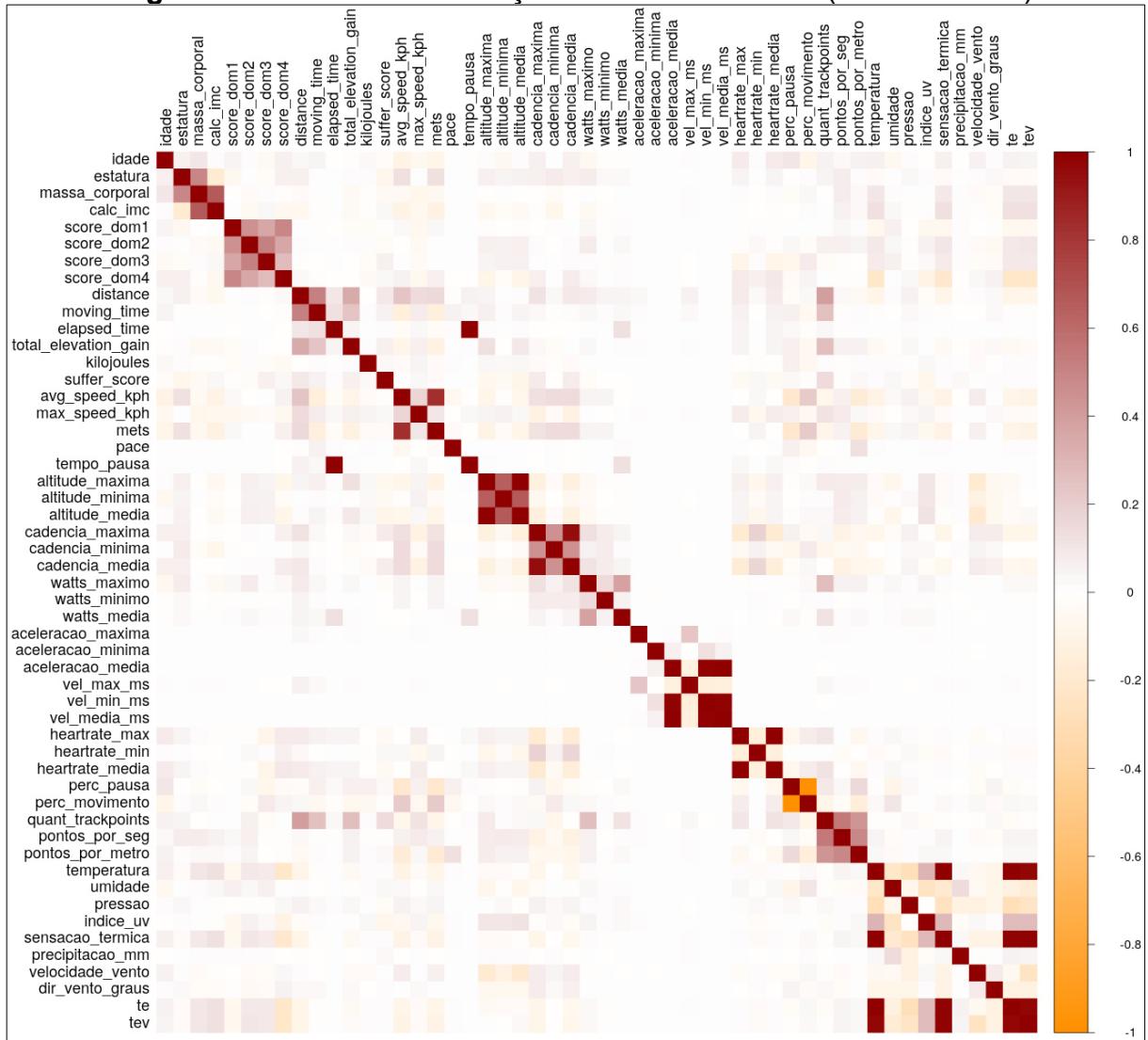
- a) distância
- b) tempo em movimento
- c) tempo decorrido
- d) velocidade máxima
- e) velocidade média
- f) ganho de elevação
- g) frequência cardíaca máxima
- h) frequência cardíaca média

Fonte: Elaborado pelo autor

A distribuição das variáveis da amostra referente à corrida de rua apresenta valores mais dispersos em relação à média, com desvio padrão elevado, o que leva à rejeição da hipótese de normalidade.

Em uma análise visual das correlações (Figura 36), observa-se com clareza a correlação somente de variáveis do mesmo grupo, o que, de novo, justifica a utilização da Mineração de Dados para buscar relações entre as mais diversas variáveis estudadas.

Figura 36 – Matriz de correlações entre as variáveis (corrida de rua)

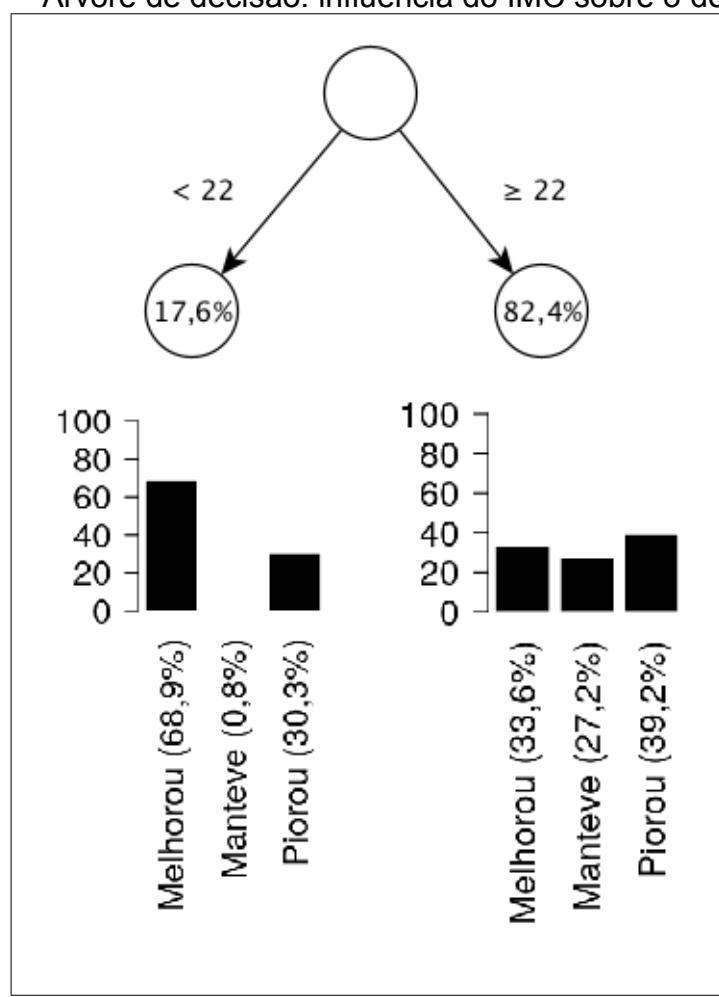


Fonte: Elaborado pelo autor

5.2.3 Mineração de Dados aplicada à corrida de rua

Analizando a influência do Índice de Massa Corporal (IMC) sobre o desempenho de corredores de rua (Figura 37), observa-se uma melhora no desempenho de forma mais significativa (68,9%) quando o IMC ficou abaixo de 22 kg/m². Assim como observado para os ciclistas, é interessante a manutenção de valores mais baixos de IMC quando se busca manter ou melhorar tal desempenho.

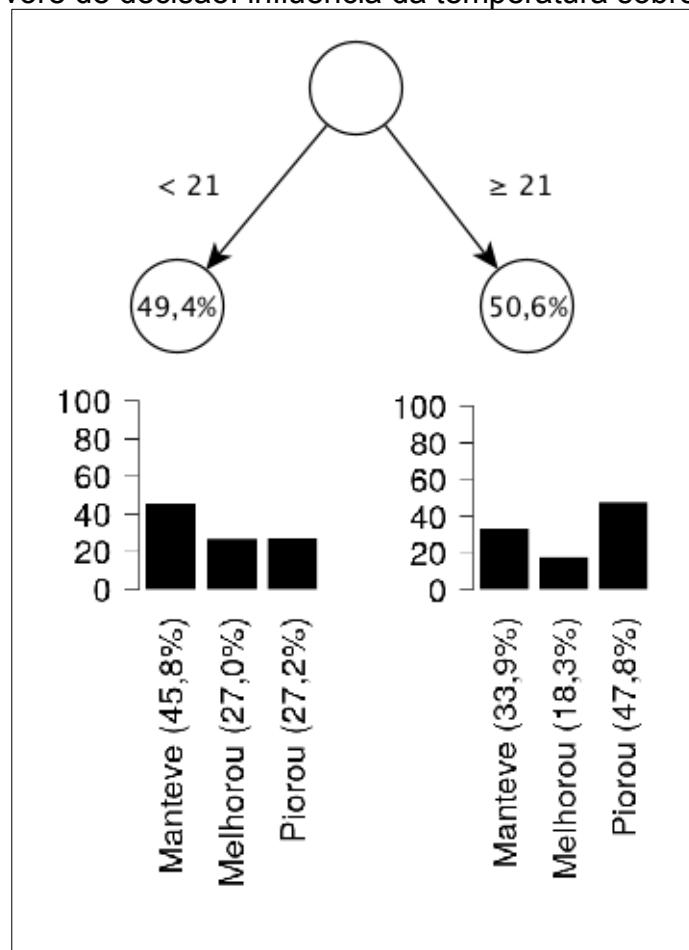
Figura 37 – Árvore de decisão: influência do IMC sobre o desempenho



Fonte: Elaborado pelo autor

No que diz respeito à influência da temperatura sobre o desempenho de corredores de rua (Figura 38), observa-se que tal desempenho piorou de forma mais significativa (47,8%) quando a temperatura ficou igual ou superior a 21º C. Para temperaturas abaixo de 21º C, observa-se de forma mais significativa a manutenção do desempenho (45,8%), ou mesmo uma melhora (27%).

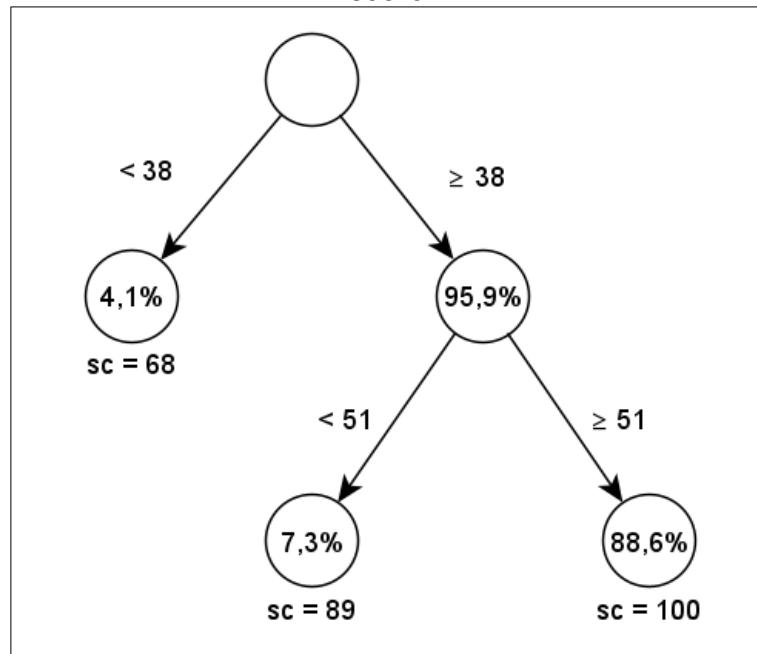
Figura 38 – Árvore de decisão: influência da temperatura sobre o desempenho



Fonte: Elaborado pelo autor

Na sequência, foi analisada a influência da umidade relativa do ar sobre o *suffer score* (sc) de corredores de rua (Figura 39), observando-se que, quando aferida acima de 51%, contribui significativamente para o aumento do *suffer score* (100), enquanto valores abaixo de 38% sinalizam para um menor *suffer score* (68).

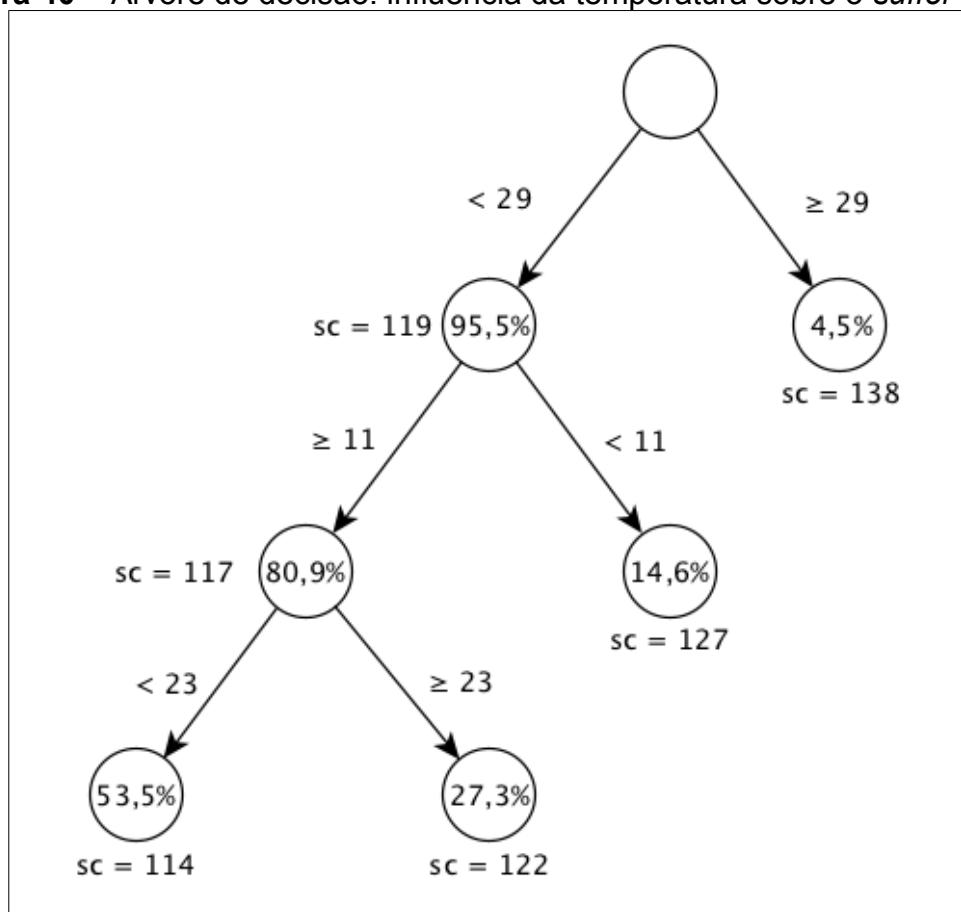
Figura 39 – Árvore de decisão: influência da umidade relativa do ar sobre o *suffer score*



Fonte: Elaborado pelo autor

O *suffer score* (sc) de corredores de rua é também influenciado pela temperatura (Figura 40). Assim, observa-se que o *suffer score* é maior tanto para temperaturas mais altas quanto para temperaturas mais baixas. Os maiores valores de *suffer score* são observados quando a temperatura ficou abaixo de 11° C (127) e quando acima de 29° C (138). Para as temperaturas acima de 11° C, observa-se um aumento do *suffer score* de forma proporcional ao aumento da temperatura.

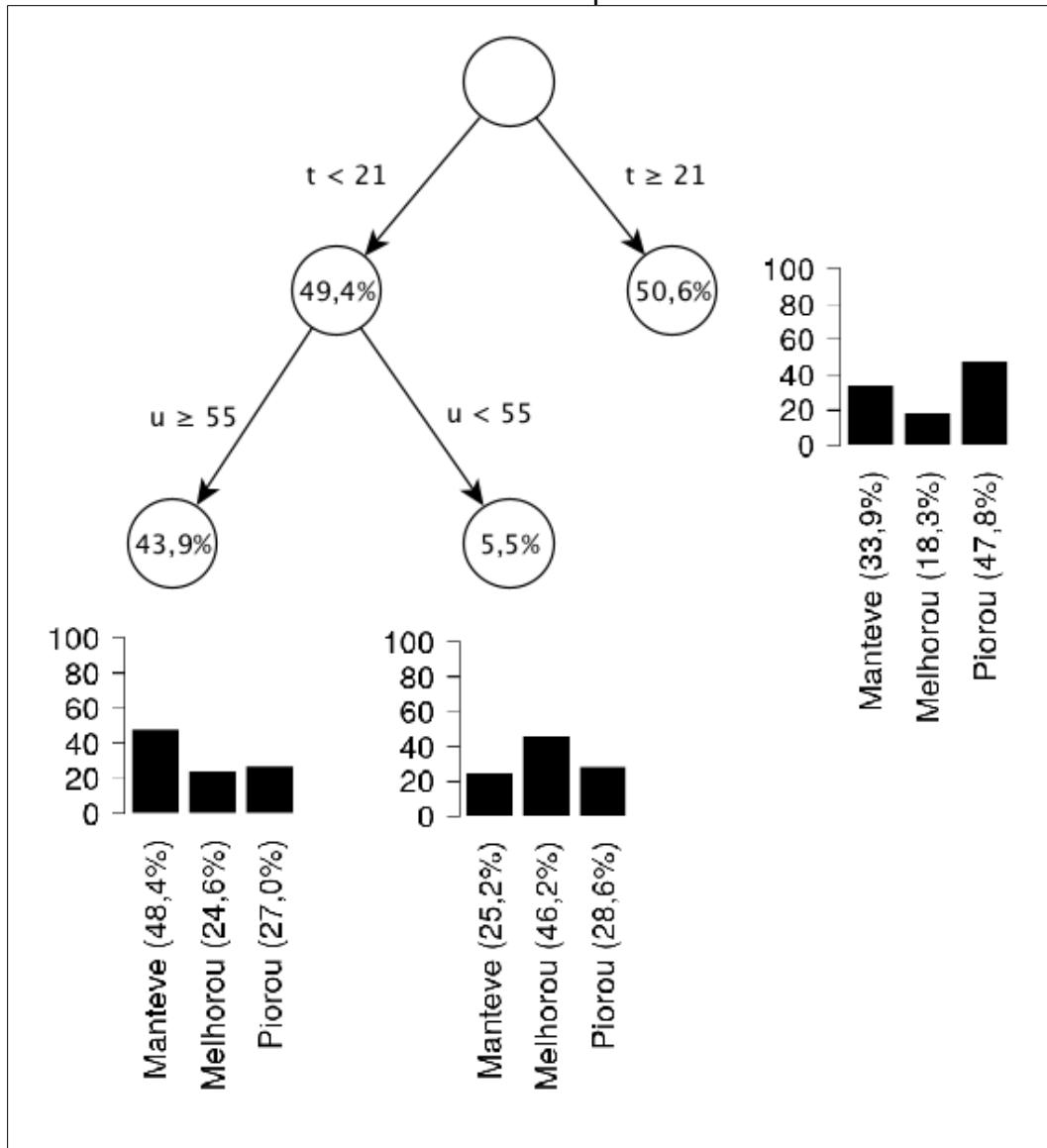
Figura 40 – Árvore de decisão: influência da temperatura sobre o suffer score



Fonte: Elaborado pelo autor

Ainda analisando a influência de variáveis climáticas sobre o desempenho de corredores de rua (Figura 41), observa-se que houve um ganho de desempenho (46,2 %) quando a temperatura (t) ficou abaixo de 21° C e a umidade relativa do ar (u) ficou abaixo de 55%. Isso sugere que a combinação de temperatura e umidade relativa do ar, ambas elevadas, condição climática característica em algumas regiões especialmente no verão, pode influenciar de forma significativa o desempenho do atleta.

Figura 41 – Árvore de decisão: influência da temperatura e umidade relativa do ar sobre o desempenho

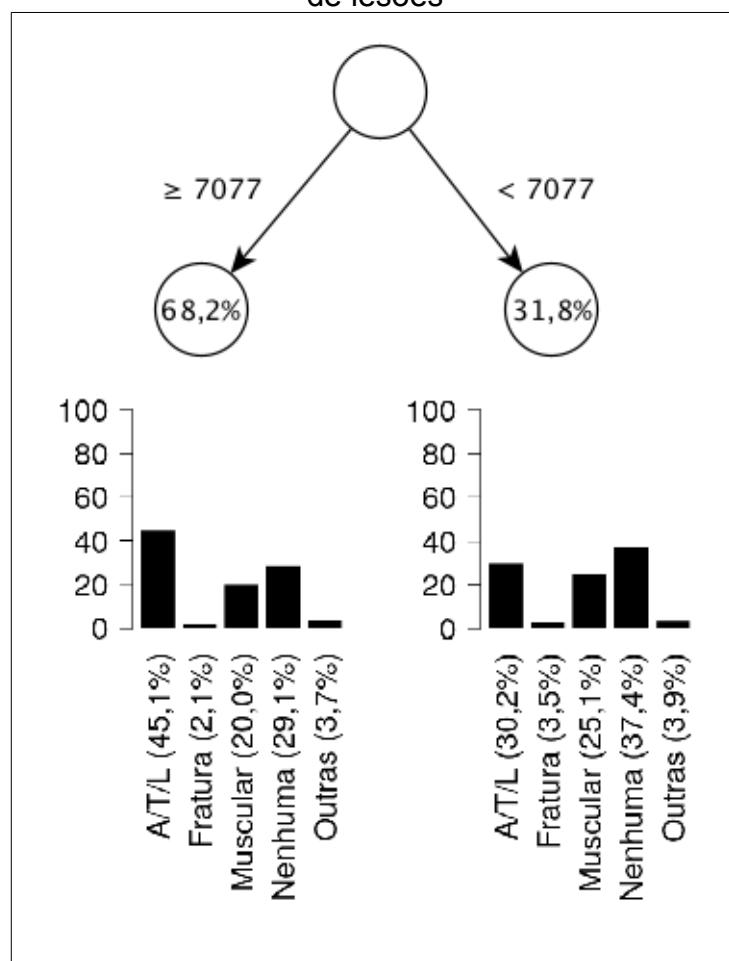


Fonte: Elaborado pelo autor

Prosseguindo, foi analisada a influência da distância percorrida pelos corredores de rua sobre o histórico de lesões (Figura 42). Observa-se uma maior incidência de lesões relacionadas a articulações, tendões ou ligamentos (A/T/L) (45,1%) quando a distância percorrida foi maior ou igual a 7077 metros, enquanto tal incidência cai para 30,2% para distâncias inferiores a 7077 metros. Van Gent et al. (2007), em uma revisão da literatura, aponta fortes evidências de que uma maior distância de treinamento por semana é um fator de risco para lesões e, como se

trata de um parâmetro modificável, tal distância não deve, preferencialmente, exceder 64 km por semana.

Figura 42 – Árvore de decisão: influência da distância percorrida sobre o histórico de lesões

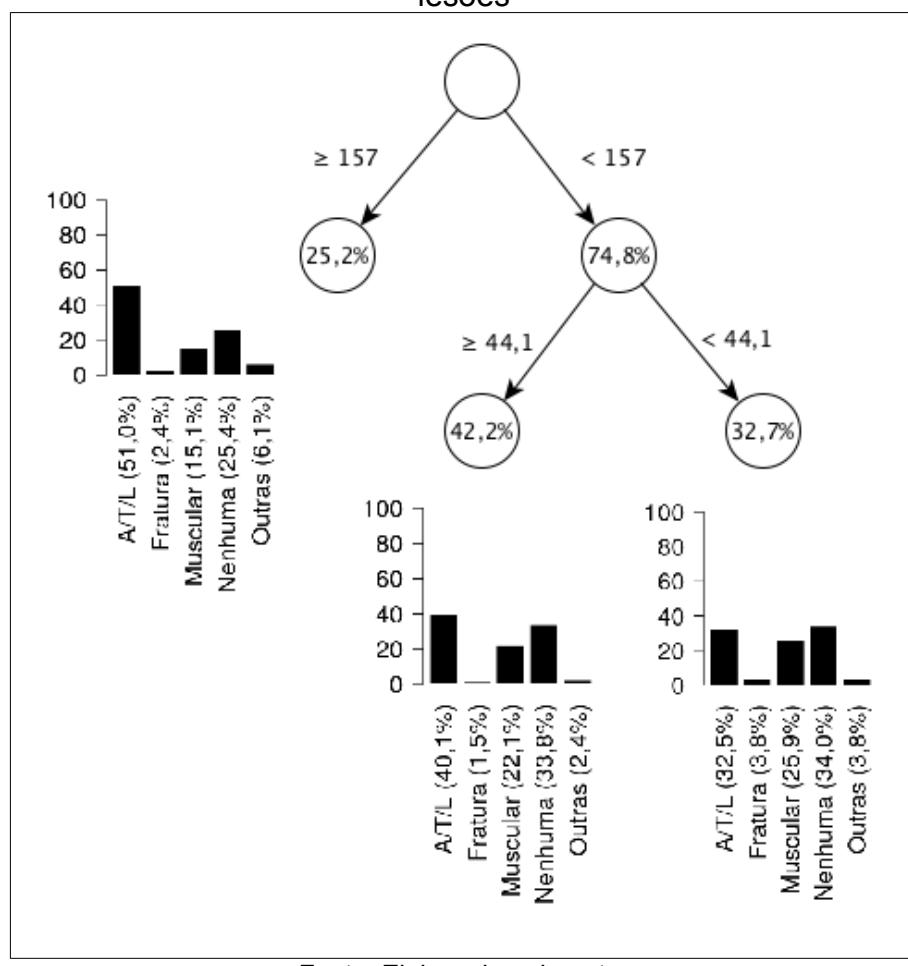


Fonte: Elaborado pelo autor

Vários estudos mostram a relação entre distância de corrida semanal e benefícios ou prejuízos à saúde (MARTI; ABELIN; MINDER, 1988; VAN GENT et al., 2007; WILLIAMS, 1997). A distância percorrida está associada diretamente ao volume, uma das variáveis fundamentais para o correto planejamento do treinamento. O aumento da distância percorrida deve ocorrer de forma gradual e equilibrada, de acordo com a intensidade do treino, o que reforça a importância da prática da corrida de rua com o devido acompanhamento profissional.

Quanto à influência do ganho de elevação sobre o histórico de lesões na corrida de rua (Figura 43), observa-se o aumento da incidência de lesões relacionadas a articulações, tendões e ligamentos (A/T/L) proporcionalmente ao aumento do ganho de elevação. Quando o ganho de elevação ficou acima de 44 metros, observa-se 40,1% de lesões do tipo A/T/L, e, quando o ganho de elevação ficou acima de 157 metros, a incidência aumenta para 51% para aquele tipo de lesão.

Figura 43 – Árvore de decisão: influência do ganho de elevação sobre o histórico de lesões

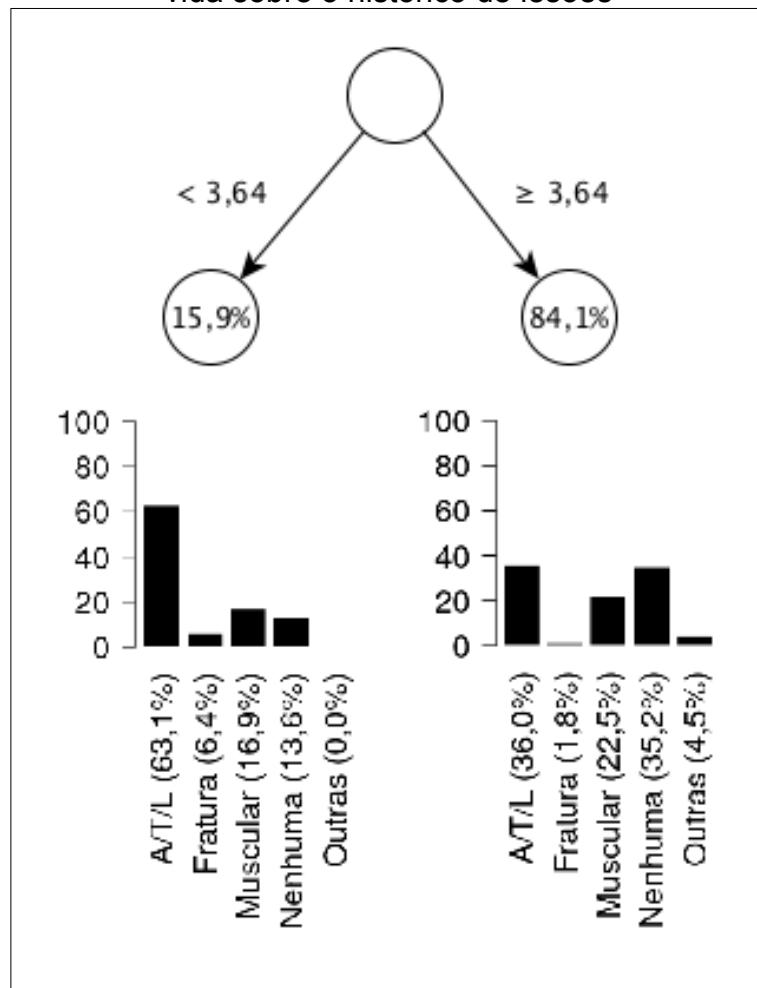


Fonte: Elaborado pelo autor

Quanto à qualidade de vida, ao analisar o *score* de corredores de rua, referente ao domínio físico (Figura 44) avaliado pelo questionário WHOQOL Abreviado, observa-se uma incidência significativamente maior de lesões do tipo A/T/L (63,1%) quando este *score* ficou abaixo de 3,64. Quando o *score* ficou acima

de 3,64 é significativamente maior o número de atletas que não reportam nenhum tipo de lesão (35,2% contra 13,6%).

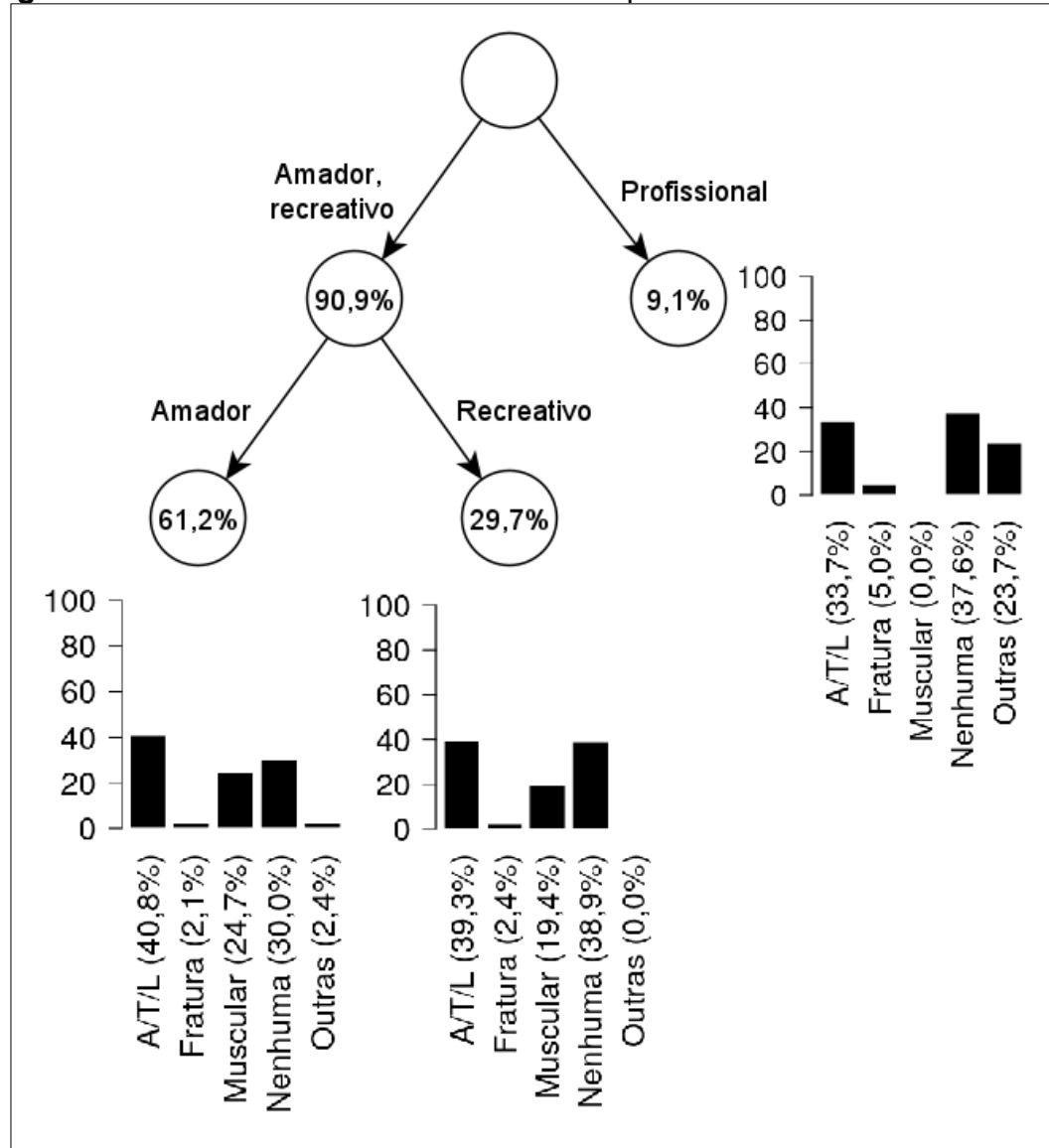
Figura 44 – Árvore de decisão: influência do score do domínio físico de qualidade de vida sobre o histórico de lesões



Fonte: Elaborado pelo autor

A influência do perfil do praticante de corrida de rua sobre o histórico de lesões (Figura 45) traz um resultado diferente do que ocorre no ciclismo. As lesões de articulação, tendão ou ligamento (A/T/L) são mais frequentes naqueles com perfil amador (40,8%) ou recreativo (39,3%). Para o grupo de profissionais, este tipo de lesão é um pouco menor (33,7%). No entanto, os profissionais relatam também a incidência de outros tipos de lesão (23,7%), bem acima das outras duas categorias. Os resultados apresentados permitem inferir que a prática de corrida de rua oferece maior risco de lesões em relação à prática do ciclismo, independentemente do perfil do praticante (recreativo, amador ou profissional).

Figura 45 – Árvore de decisão: influência do perfil sobre o histórico de lesões

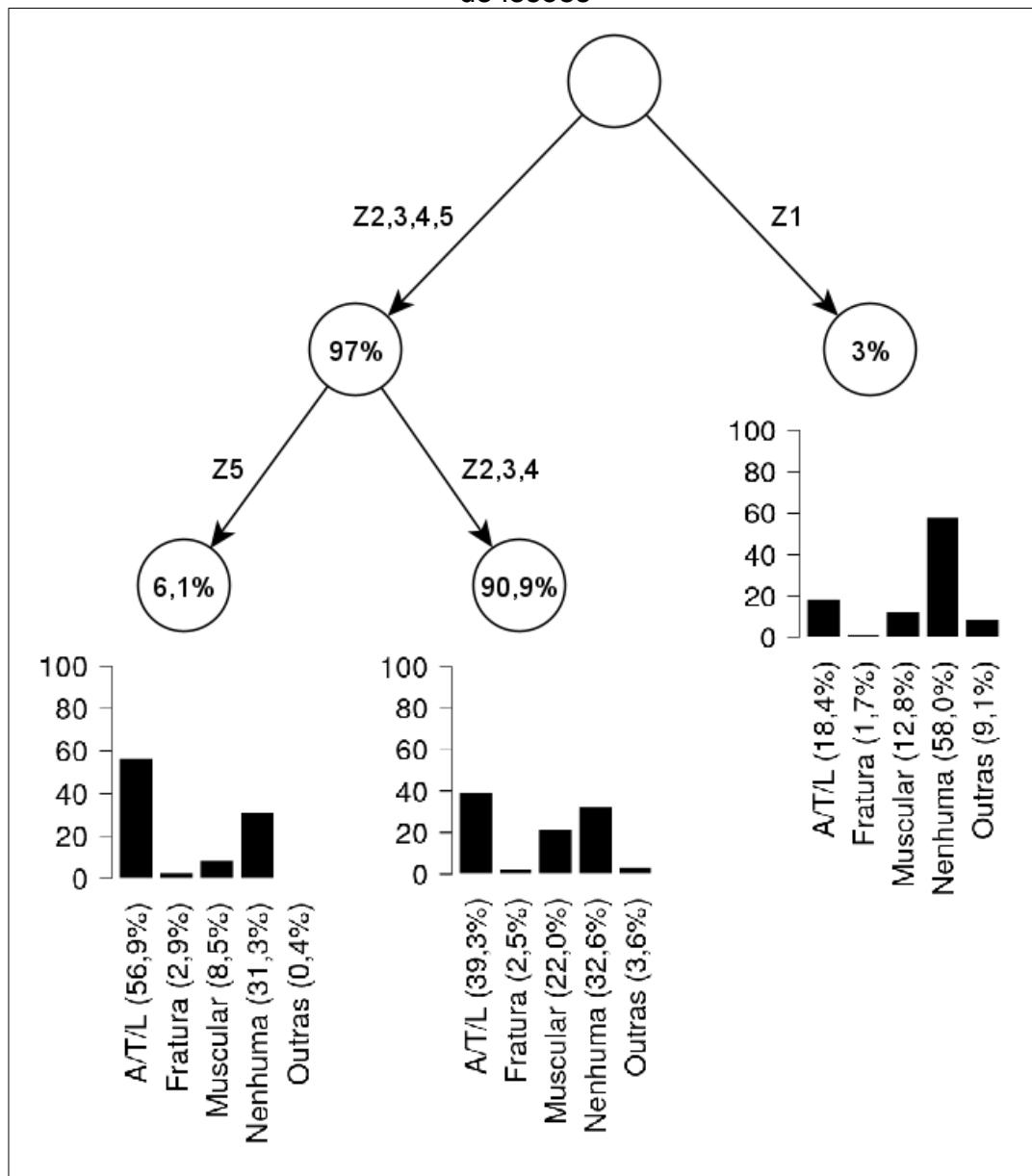


Fonte: Elaborado pelo autor

Quanto à influência das zonas de intensidade predominantes durante a prática da corrida de rua sobre o histórico de lesões (Figura 46), observa-se que a maior porcentagem de incidência de lesões do tipo A/T/L (56,9%) está associada ao nível mais elevado de intensidade (Z5). Ainda, as lesões musculares são mais frequentes nas zonas intermediárias de intensidade (Z2, Z3 e Z4), com incidência de 32,6%, enquanto encontramos valores bem baixos na zona 1 (12,8%) e na zona 5 (8,5%). Mesmo na zona de intensidade mais leve (zona 1), a ocorrência de lesões é significativa (18,4% para lesões A/T/L, 12,8% para lesões musculares e 9,1% para outros tipos), embora o quantidade de praticantes que não reportam

nenhum tipo de lesão seja bem maior quando comparado às demais zonas de intensidade (58% contra aproximadamente 32% das outras zonas).

Figura 46 – Árvore de decisão: influência das zonas de intensidade sobre o histórico de lesões



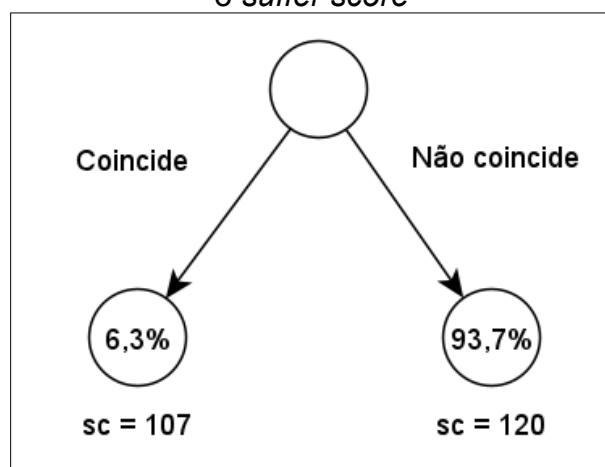
Fonte: Elaborado pelo autor

Deve-se observar que boa parte dos problemas relacionados à corrida de rua podem ser evitados por meio de uma adaptação adequada ao treinamento e por meio do incremento dosado de forma criteriosa tanto do volume quanto da intensidade do treinamento (FREDERICSON; MISRA, 2007), bem como pela

observação das variáveis envolvidas, como, por exemplo, fatores climáticos e condição física.

Analizando a influência da direção do vento comparada com a direção do percurso sobre o impacto no esforço realizado pelo praticante de corrida de rua (Figura 47), observou-se, por meio do indicador *suffer score*, um maior esforço quando a direção do vento não coincide com a direção do percurso ($sc = 120$ contra $sc = 107$ quando as direções são coincidentes).

Figura 47 – Árvore de decisão: influência das direções do vento e do percurso sobre o *suffer score*



Diferente do ciclismo, não observou-se na corrida de rua um impacto da direção do vento no desempenho ou na velocidade máxima obtida. No entanto, observou-se um maior esforço quando a direção do vento não coincide com a direção do percurso.

5.3 Considerações finais sobre o estudo experimental

O estudo experimental apresentado neste capítulo traz elementos suficientes para mostrar a viabilidade de se utilizar técnicas da Mineração de Dados na busca de conhecimentos que possam auxiliar na tomada de decisões referentes ao treinamento físico e esportivo. Para tanto, foi construído o protótipo de uma plataforma computacional voltada para implementar algoritmos da Mineração de

Dados, especificamente Árvores de Decisão, de forma que possam ser aplicados na rotina diária do treinamento de atletas de ciclismo e corrida de rua, tanto profissionais quanto amadores, treinados e supervisionados por profissionais de Educação Física. O próximo capítulo aborda a construção desta plataforma.

6 PLATAFORMA *FITNESS TOOLS*

A plataforma computacional proposta no presente estudo, denominada *Fitness Tools*, é o protótipo de uma aplicação disponível via Internet³⁶, por meio da qual o profissional de Educação Física é capaz de realizar o cadastro de seus clientes e o registro de dados, bem como análises referentes às atividades desenvolvidas. A partir dos dados registrados, o profissional passa a contar com ferramentas de análise por meio de algoritmos da Mineração de Dados, capazes de fornecer informações que possam auxiliá-lo no planejamento, acompanhamento e ajustes relacionados ao treinamento esportivo.

Este capítulo apresenta as informações gerais sobre a plataforma, contemplando sua construção, questões referentes à segurança e privacidade, entrada e análise de dados. Informações mais detalhadas e especificações técnicas estão disponíveis no Apêndice I.

6.1 Construção da plataforma

A plataforma *Fitness Tools* é uma aplicação disponível via Internet, por meio do servidor web *Apache*³⁷. Devido a sua característica multiplataforma, pode ser acessada por meio de diferentes dispositivos, tais como *smartphone*, computador de mesa, tablet e notebook, além de diferentes sistemas operacionais, como *Windows*, *MacOS*, *Linux*, *Android* e *iOS*.

A plataforma foi construída por meio do *framework*³⁸ *Flask*³⁹, baseado na linguagem de programação *Python*. Para armazenamento e recuperação dos dados foi utilizado o sistema de gerenciamento de bancos de dados *MySQL Community Server* versão 5.7.26. O banco de dados foi construído conforme o Modelo Entidade Relacionamento apresentado no Apêndice J.

36 Disponível no endereço eletrônico: <https://www.fitnesstools.com.br:5100/app>

37 Disponível em <https://httpd.apache.org/>

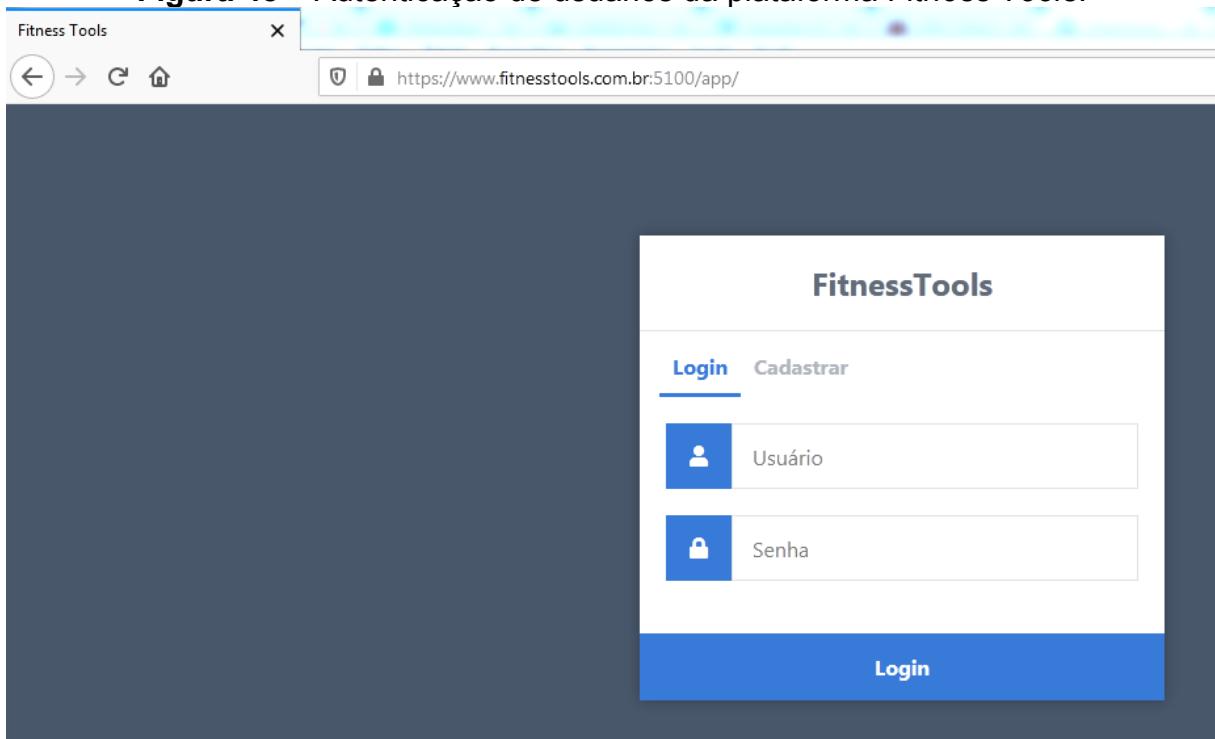
38 *Framework*, no contexto de desenvolvimento de software, consiste em um conjunto de implementações pré-programadas e reutilizáveis, ou seja, uma plataforma base para desenvolvimento de aplicações (RIEHLE, 2000).

39 Disponível em <https://palletsprojects.com/>

6.2 Segurança e privacidade

O acesso à plataforma *Fitness Tools* é realizado por meio de autenticação de usuários cadastrados, com a respectiva senha (Figura 48).

Figura 48 – Autenticação de usuários da plataforma *Fitness Tools*.



Fonte: Elaborado pelo autor

A plataforma conta com segurança no tráfego de dados por meio de um canal criptografado entre o servidor web e o software navegador de Internet do usuário, proporcionando sigilo e segurança. O servidor conta com *firewall*, permitindo ou bloqueando os tráfegos específicos, de acordo com um conjunto de regras de segurança pré-definidas. Por fim, são realizadas cópias de segurança (*backup*) diariamente.

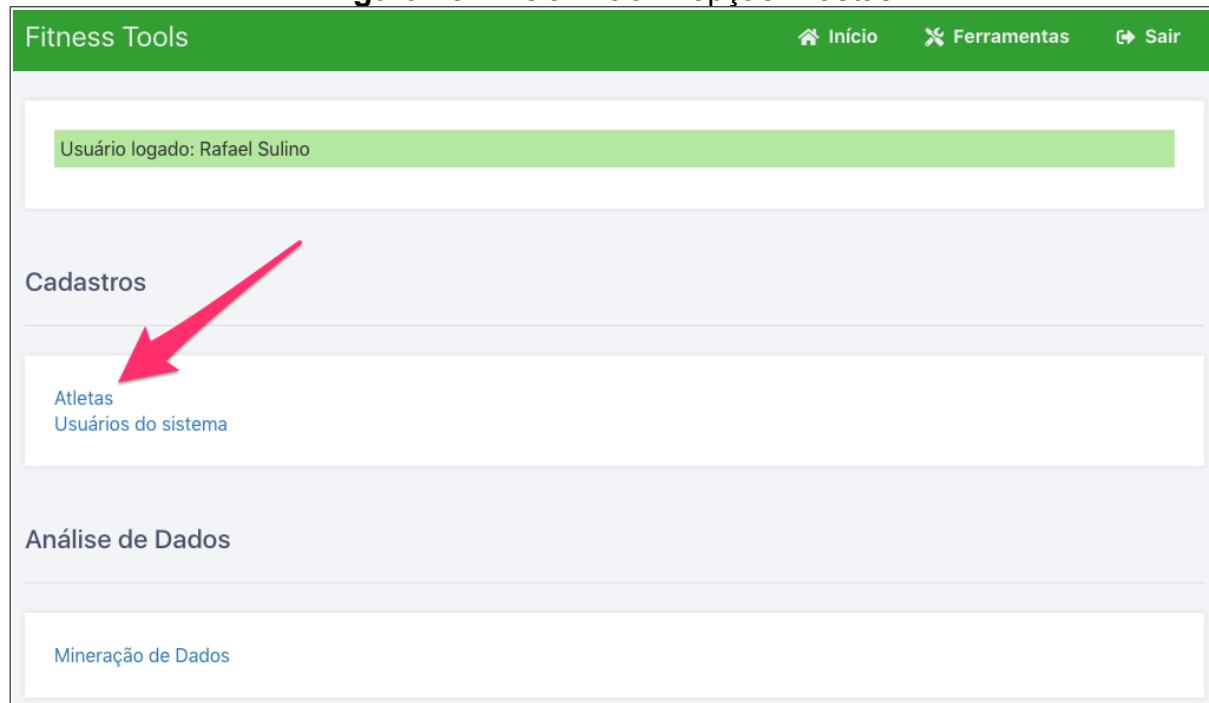
6.3 Entrada de dados

A plataforma possibilita o registro de dados referentes aos usuários (profissionais), atletas e respectivas atividades realizadas, a partir de diferentes

fontes. Os dados básicos dos usuários e atletas são registrados por meio de formulários de cadastro e anamnese⁴⁰, e os referentes à qualidade de vida usando formulário baseado na versão abreviada do questionário WHOQOL (FLECK et al., 2000). Já os dados referentes às atividades realizadas e aos climáticos são inseridos na plataforma de forma automática por meio da API do *Strava* e da API do serviço *World Weather Online*, respectivamente. Por fim, a plataforma possibilita o registro de ocorrências relacionadas aos atletas e atividades realizadas (lesões, desconforto, entre outros eventos), por meio de formulário próprio.

O acesso a todas as opções para a entrada de dados referentes aos atletas é feito por meio da opção “Atletas” na tela inicial da plataforma (Figura 49).

Figura 49 – Tela inicial – opção “Atletas”



Fonte: Elaborado pelo autor

O cadastro de usuários da plataforma é feito por meio de solicitação prévia, via e-mail, conforme instruções apresentadas na página inicial da plataforma. Quanto ao cadastro dos atletas, parte dos dados é inserida de forma manual pelo

⁴⁰ No contexto da área da Saúde, trata-se de um conjunto de informações sobre o histórico de saúde do indivíduo, geralmente conduzido na forma de entrevista ou questionário (BARBANTI, 2003).

usuário e outra parte inserida de forma automática por meio da integração com a API do *Strava*.

As informações referentes à anamnese são inseridas por meio da aba “Anamnese”. O registro das informações consiste em um questionário que abrange diversos tópicos, como histórico de lesões, uso de medicamentos, histórico de diabetes, hipertensão, entre outras informações.

O registro de dados referente a qualidade de vida é feito por meio da aba “Qualidade de vida”, contendo as 26 perguntas do instrumento WHOQOL abreviado (Anexo C).

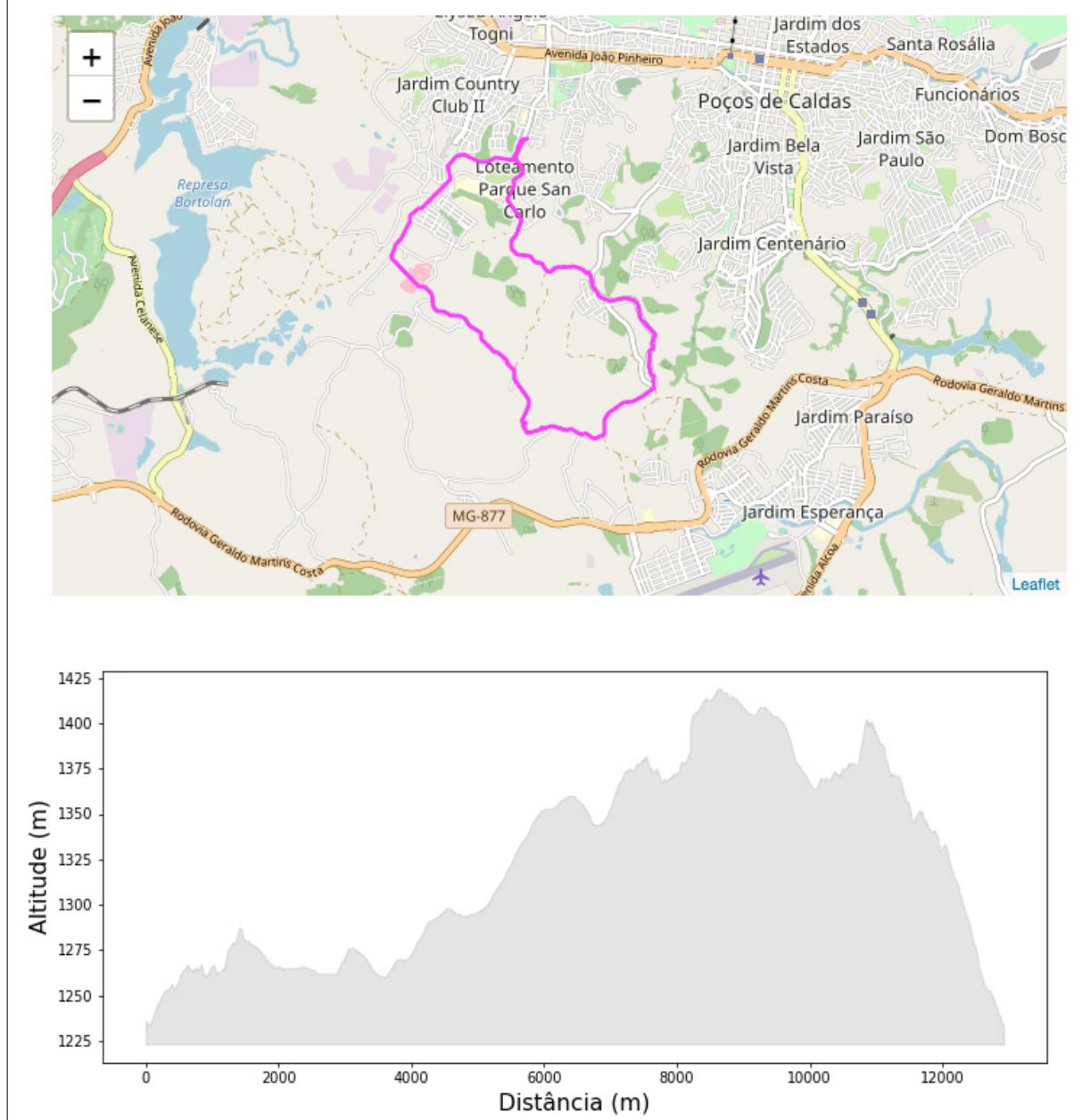
A plataforma conta com a integração de dados externos, tanto das atividades realizadas quanto climáticos. Os dados referentes às atividades praticadas podem ser obtidos diretamente por meio da API do *Strava*. Já os dados climáticos são integrados através da API do serviço *World Weather Online*.

6.4 Análise de dados

A plataforma oferece ao usuário recursos para análise de dados tanto de atividades individuais quanto do conjunto de dados como um todo, por meio da Mineração de Dados.

Em um primeiro momento, as atividades registradas na plataforma podem ser pré-analisadas individualmente, sem o uso da Mineração de Dados. A análise individual apresenta diversas informações referentes ao percurso e desempenho do atleta. A Figura 50 destaca o mapa do percurso bem como sua altimetria, com base nas informações de latitude, longitude e altitude, especificamente para a atividade analisada.

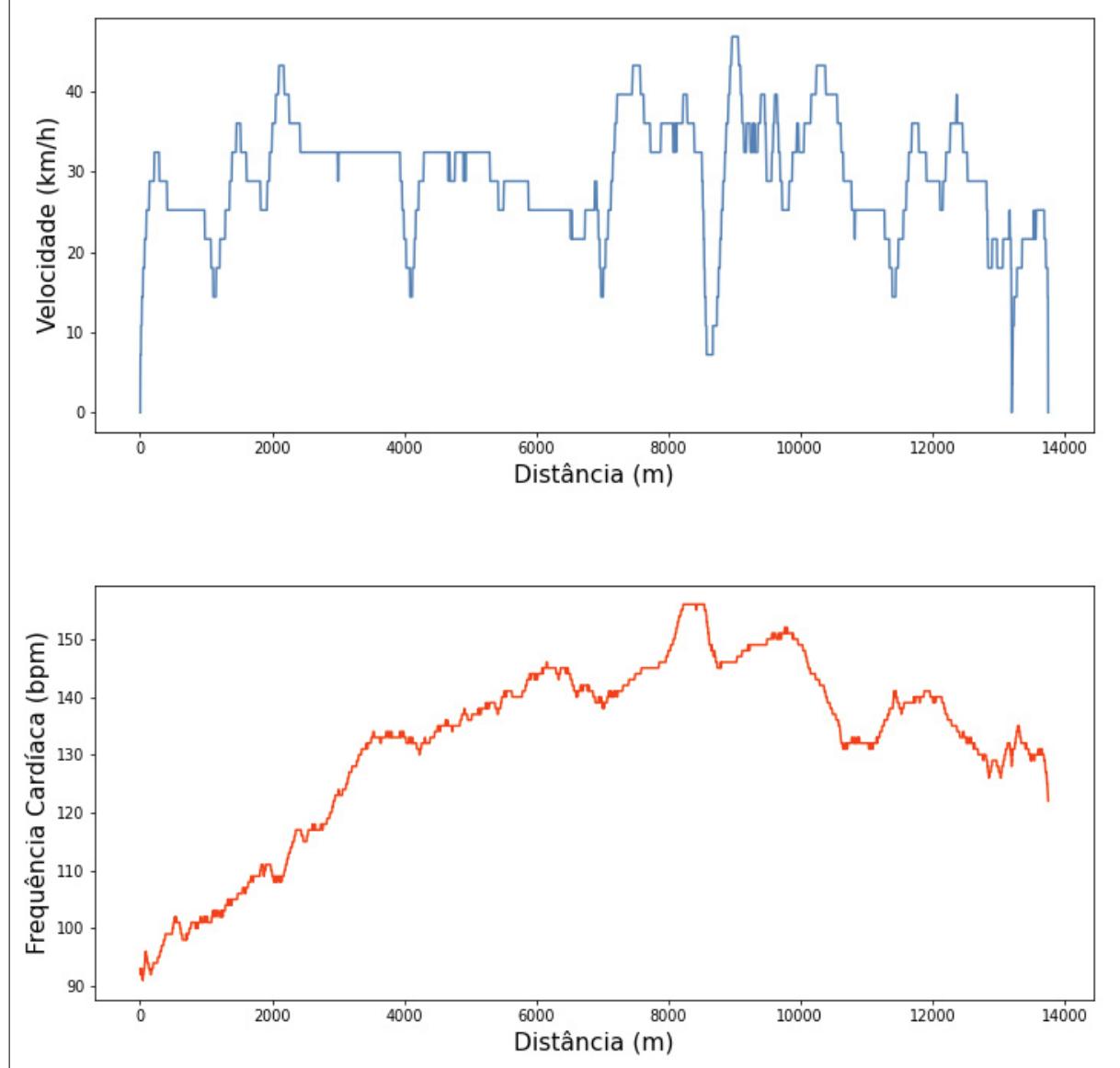
Figura 50 – Exemplo de análise individual (mapa e altimetria)



Fonte: Elaborado pelo autor

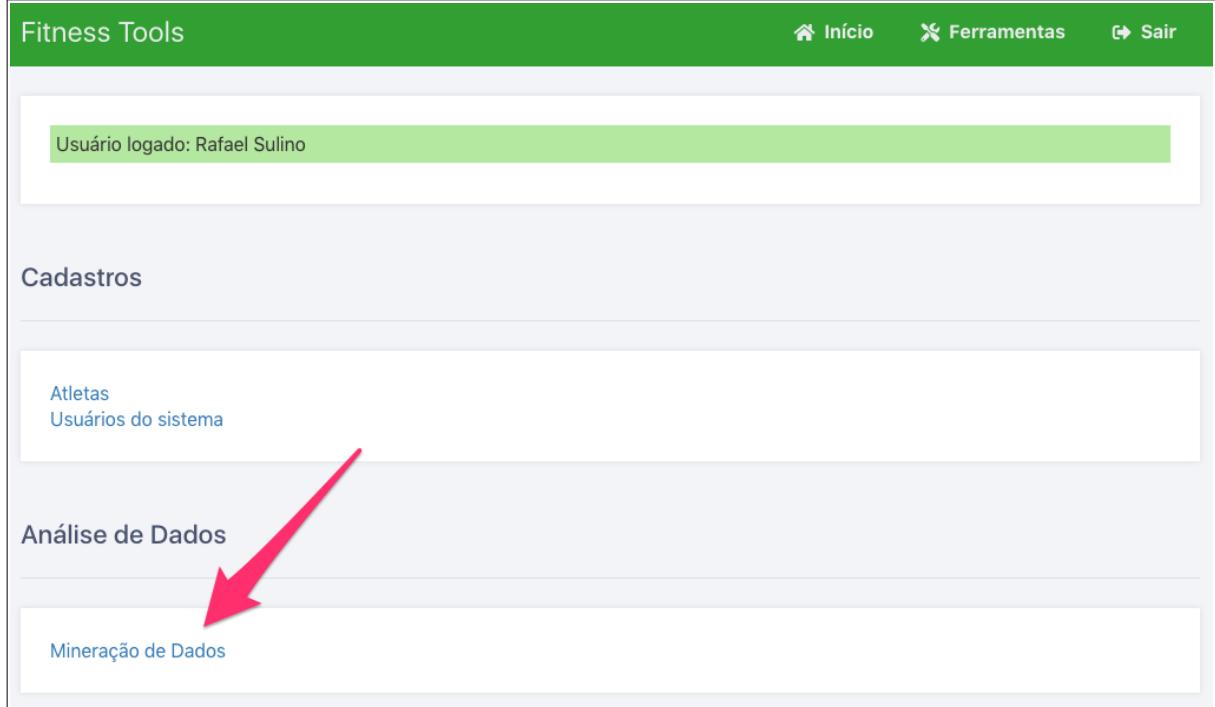
A Figura 51 destaca a velocidade e a frequência cardíaca durante o percurso.

Figura 51 – Exemplo de análise individual (velocidade e frequência cardíaca)



Fonte: Elaborado pelo autor

O acesso para a ferramenta de análise de dados por meio da Mineração de Dados é realizada usando a opção “Mineração de Dados” na tela inicial da plataforma *Fitness Tools* (Figura 52).

Figura 52 – Tela inicial – opção “Mineração de Dados”

Fonte: Elaborado pelo autor

Para uso da Mineração de Dados (Figura 53), é preciso escolher o tipo de atividade, disponível, a princípio, somente para ciclismo e corrida de rua, e a forma da análise desejada, com base nas análises relevantes obtidas por meio do estudo experimental, além da seleção do número máximo de ramificações da árvore de decisão.

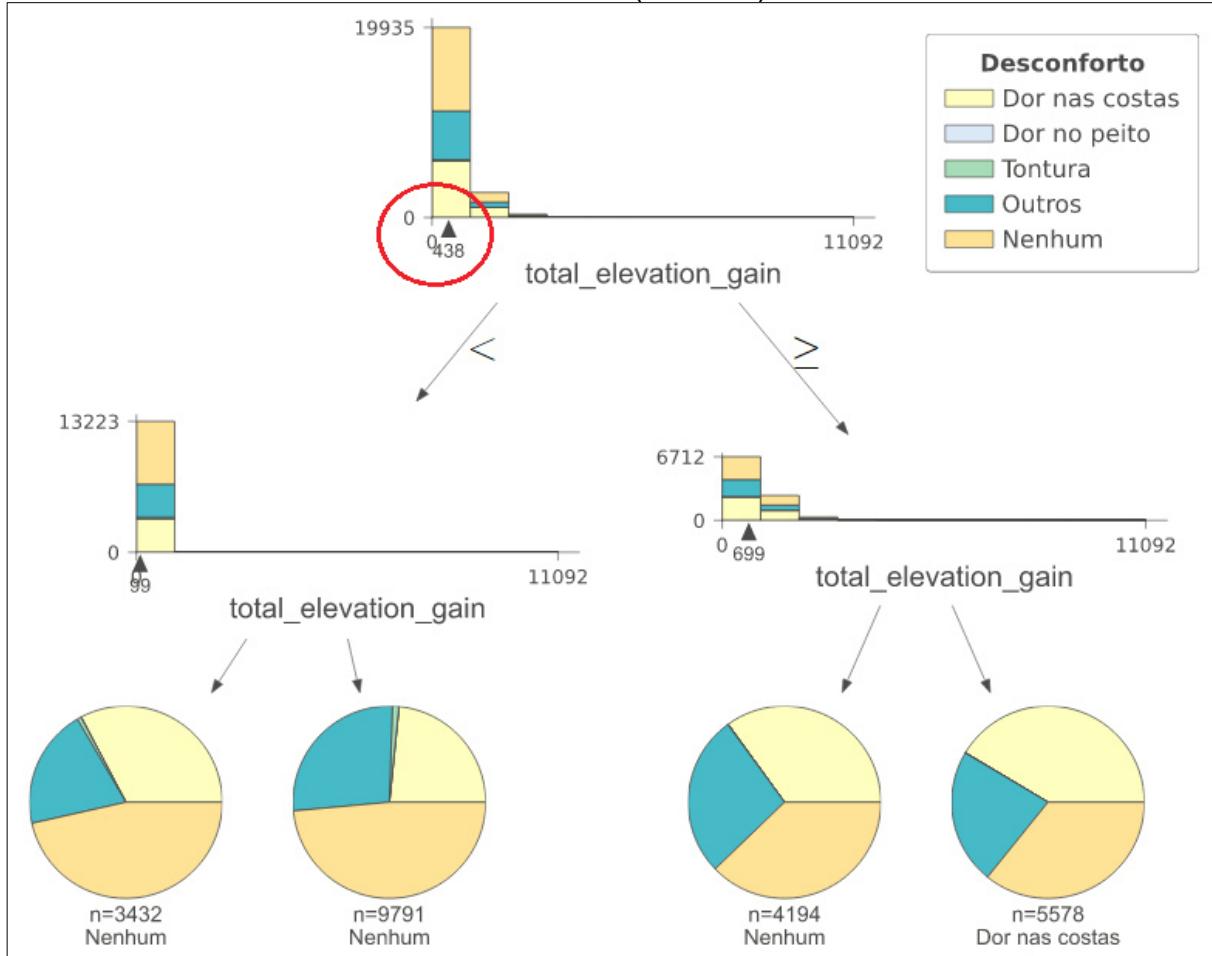
Figura 53 – Ferramenta de Mineração de Dados

The screenshot shows a web-based application titled "Fitness Tools". At the top, there is a green header bar with the text "Fitness Tools" on the left and navigation links "Início", "Ferramentas", and "Sair" on the right. Below the header, a green banner displays the text "Usuário logado:". The main content area is titled "Mineração de Dados". It contains three input fields: "Tipo de atividade:" with a dropdown menu set to "Ciclismo", "Escolha a análise desejada:" with a dropdown menu set to "Intensidade x Lesões", and "Níveis de ramificação da árvore (padrão=2):" with a value of "2" in a text input field. A "Processar" button is located at the bottom of this section.

Fonte: Elaborado pelo autor

O resultado do processamento usando a Mineração de Dados é apresentado de forma gráfica por meio de uma Árvore de Decisão, composta por gráficos detalhados (Figura 54). Este exemplo mostra uma análise da influência do ganho de elevação sobre a percepção de desconforto nas atividades de ciclismo. Quando este ganho de elevação ficou igual ou maior que 438 metros, observa-se uma maior incidência na percepção de dor nas costas, representada nos gráficos inferiores à direita, pela cor amarela mais clara.

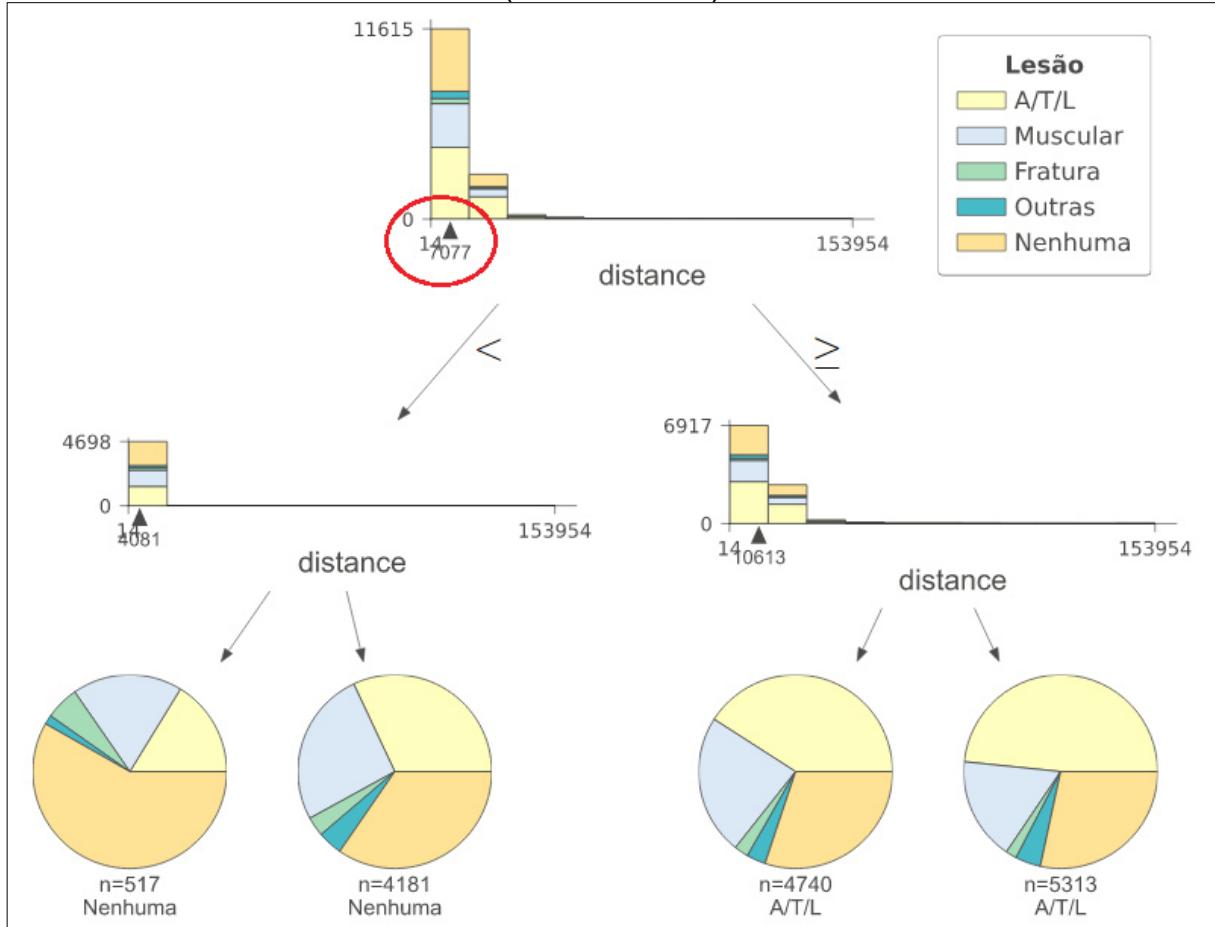
Figura 54 – Exemplo 1: influência do ganho de elevação na percepção de desconforto (ciclismo)



Fonte: Elaborado pelo autor

Em um segundo exemplo (Figura 55), podemos analisar a influência da distância percorrida sobre o histórico de lesões em atividades de corrida de rua. Observa-se uma maior incidência de lesões relacionadas a articulações, tendões ou ligamentos (A/T/L) quando a distância percorrida foi maior ou igual a 7077 metros (representado nos gráficos inferiores à direita pela cor amarela mais clara).

Figura 55 – Exemplo 2: influência da distância percorrida sobre o histórico de lesões (corrida de rua)



Fonte: Elaborado pelo autor

As informações obtidas por meio das análises usando Mineração de Dados podem agregar valor ao treinamento esportivo, principalmente para o profissional, pois os resultados possibilitam alertar para determinados cuidados, além de direcionar o foco dos ajustes para o treinamento de seus clientes.

7 CONSIDERAÇÕES FINAIS

A rotina diária do profissional de Educação Física que atua diretamente com o treinamento esportivo envolve avaliações, acompanhamento de resultados, prescrição de treinamento e ajustes no treinamento em andamento, para cada um de seus alunos ou clientes. De modo geral, a tecnologia, especificamente aquela voltada para análise de dados, ainda é muito limitada no dia a dia de muitos profissionais. Na maioria das vezes, os ajustes na rotina de treinamento do atleta são realizados levando-se em consideração apenas os relatos e percepções do próprio atleta.

O estudo apresentado mostra o potencial de utilização da Mineração de Dados na descoberta de conhecimentos que possam ser utilizados para a tomada de decisões quanto à prática de atividade física de forma mais segura e eficiente. A Mineração de Dados oferece uma opção flexível, adaptável a características específicas (região, clima, tipo de terreno), podendo ser utilizada para um único atleta, por meio de ajustes levando em conta seu próprio histórico (como volume, intensidade, características do terreno e clima), ou para grandes grupos de atletas, como os populares grupos de corrida. O ajuste adequado do volume e da intensidade do treinamento seria útil, por exemplo, para a prevenção de lesões.

Embora não seja o foco desta pesquisa, o uso da Mineração de Dados possibilitou, inclusive, fazer algumas comparações para determinados conjuntos de variáveis, especificamente para as duas modalidades analisadas (ciclismo e corrida de rua), e inferir, por exemplo, que a prática de corrida de rua oferece maior risco de lesões, independentemente do perfil do praticante (recreativo, amador ou profissional). Conhecendo seu cliente, após uma anamnese e outras avaliações, e valendo-se de tais comparações e inferências, o profissional poderá, por exemplo, decidir qual atividade seria, a princípio, mais indicada.

Atualmente, é plenamente possível aliar a farta disponibilidade de dados, obtidos principalmente por meio de aplicativos gratuitos como o *Strava*, com técnicas de análise de dados para extração de conhecimento. O uso da Mineração de Dados,

especificamente na área da Educação Física, possibilita identificar os limites para determinadas variáveis em um certo contexto (por exemplo, os limites referentes ao ganho de elevação e sua relação com o desconforto do praticante, conforme apresentado no estudo).

A plataforma *Fitness Tools* apresentada neste estudo traz uma contribuição que pode ser importante para a área da Educação Física, oferecendo aos profissionais acesso, de forma simples e intuitiva, a tecnologias avançadas de análise de dados. Por meio dos recursos disponíveis na plataforma, o profissional contará com informações que o auxiliem na tomada de decisões, no ajuste fino do treinamento de seus clientes e na prevenção de ocorrências que possam prejudicar o desempenho do atleta, como lesões e desconforto. Em um primeiro momento, tais recursos estão disponíveis somente para atividades de ciclismo e corrida de rua. Para estas atividades, o profissional poderá, por exemplo, identificar o quanto a direção do vento influencia no desempenho do atleta, facilitando a escolha de um percurso mais adequado, com o auxílio de informações sobre previsão do tempo para a data e hora do treino planejado, ou preparar o atleta para as condições específicas de uma competição, baseado em informações sobre o percurso e o clima.

REFERÊNCIAS

- AGGARWAL, C. C. **Data Mining: The Textbook**. New York: Springer International Publishing, 2015.
- AMARAL, F. **Introdução à ciência de dados: mineração de dados e Big Data**. Rio de Janeiro: Alta Books, 2016.
- AMERICAN COLLEGE OF SPORTS MEDICINE. **ACSM's guidelines for exercise testing and prescription**. 9. ed. Philadelphia: Lippincott Williams & Wilkins, 2013.
- ANSARI, M.; NOURIAN, R.; KHODAEE, M. Mountain Biking Injuries. **Current Sports Medicine Reports**, v. 16, n. 6, p. 404–412, 2017.
- APPLE INC. **Efetuar um ECG com a app ECG no Apple Watch Series 4 ou posterior**. Disponível em: <<https://support.apple.com/pt-pt/HT208955>>. Acesso em: 7 maio. 2020.
- ARSLAN, S. S. et al. The effects of physical activity on sleep quality, job satisfaction, and quality of life in office workers. **Work**, v. 63, n. 1, p. 3–7, 1 jan. 2019.
- ASIMOV, I. **Fundação: trilogia: fundação, fundação e imperio, segunda fundação**. São Paulo: Hemus, 1975.
- ASPLUND, C.; WEBB, C.; BARKDULL, T. Neck and Back Pain in Bicycling. **Current Sports Medicine Reports**, v. 4, p. 271–274, 2005.
- AVILA, C. O. Novel Use of Apple Watch 4 to Obtain 3-Lead Electrocardiogram and Detect Cardiac Ischemia. **The Permanente Journal**, v. 23, 2019.
- BANAEE, H.; AHMED, M. U.; LOUTFI, A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. **Sensors**, v. 13, n. 12, p. 17472–17500, 2013.
- BARBANTI, V. J. **Dicionário de Educação Física e Esporte**. 2. ed. Barueri, SP: Manole, 2003.
- BENNELL, K. L.; CROSSLEY, K. Musculoskeletal injuries in track and field: incidence, distribution and risk factors. **Australian journal of science and medicine in sport**, v. 28, n. 3, p. 69–75, 1996.
- BISWAS, D. et al. Heart Rate Estimation from Wrist-Worn Photoplethysmography: A Review. **IEEE Sensors Journal**, v. 19, n. 16, p. 6560–6570, 2019.
- BIZE, R.; JOHNSON, J. A.; PLOTNIKOFF, R. C. Physical activity level and health-related quality of life in the general adult population: A systematic review. **Preventive Medicine**, v. 45, n. 6, p. 401–415, dez. 2007.

CAMERON, F. **What is NFC? Everything you need to know.** Disponível em: <<https://www.techradar.com/news/what-is-nfc>>. Acesso em: 7 set. 2020.

CASPERSEN, C. J.; POWELL, K. E.; CHRISTENSON, G. M. Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research. **Notes and Queries**, v. s9-IX, n. 228, p. 365–366, 1985.

CASTELLI. **Zwift - Ride it out with Castelli.** Disponível em: <<https://blog.castelli-cycling.com/2020/03/26/zwift-ride-it-out-with-castelli/>>. Acesso em: 7 maio. 2020.

CASTRO, L. N. D.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações.** São Paulo: Saraiva, 2016.

CHALKO, T. J. High accuracy speed measurement using GPS (Global Positioning System). **NU Journal of Discovery**, v. 4, p. 1–9, 2007.

CINTIA, P.; PAPPALARDO, L.; PEDRESCHI, D. Engine matters: a first large scale data driven study on cyclists' performance. **Proceedings - IEEE 13th International Conference on Data Mining Workshops**, p. 147–153, 2013.

COLE, E.; KRUTZ, R.; CONLEY, J. W. **Network Security Bible.** Indianapolis, IN: Wiley Publishing Inc., 2005.

CÔRTES, S. C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados: funcionalidades, técnicas e abordagens.** Rio de Janeiro: PUC, 2002.

CULOS-REED, S. N.; BRAWLEY, L. R. Fibromyalgia, physical activity, and daily functioning: The importance of efficacy and health-related quality of life. **Arthritis Care & Research**, v. 13, n. 6, p. 343–351, 2000.

DAHN, J. R.; AND, F. J. P. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. **Current Opinion in Psychiatry**, v. 18, n. 2, p. 189–193, 2005.

DANTAS, E. H. M. **A Prática da Preparação Física.** 6. ed. São Paulo: Roca, 2014.

DATE, C. J. **Introdução a sistemas de bancos de dados.** Rio de Janeiro: Elsevier Brasil, 2004.

DAVENPORT, T. H.; KLAHR, P. Managing Customer Support Knowledge. **California Management Review**, v. 40, n. 3, p. 195–208, 2012.

DE AMO, S. Técnicas de mineração de dados. **Jornada de Atualização em Informática da Universidade Federal de Uberlândia**, 2004.

DE HARTOG, J. J. et al. Do the Health Benefits of Cycling Outweigh the Risks? **Environmental Health Perspectives**, v. 118, n. 8, p. 1109–1116, ago. 2010.

DETTORI, N. J.; NORVELL, D. C. Non-traumatic bicycle injuries: A review of the literature. **Sports Medicine**, v. 36, n. 1, p. 7–18, 2006.

- DUC, S. et al. Muscular activity during uphill cycling: Effect of slope, posture, hand grip position and constrained bicycle lateral sways. **Journal of Electromyography and Kinesiology**, v. 18, n. 1, p. 116–127, 2008.
- DUTTA, S. Strategies for implementing knowledge-based systems. **IEEE Transactions on Engineering Management**, v. 44, n. 1, p. 79–90, 1997.
- ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados**. São Paulo: Pearson Addison Wesley, 2005.
- FANG, Y.-Y.; HUANG, C.-Y.; HSU, M.-C. Effectiveness of a physical activity program on weight, physical fitness, occupational stress, job satisfaction and quality of life of overweight employees in high-tech industries: a randomized controlled study. **International Journal of Occupational Safety and Ergonomics**, v. 25, n. 4, p. 621–629, 2 out. 2019.
- FANTONI, A. Dispositivos wearable para o campo da saúde: reflexões acerca do monitoramento de dados do corpo humano. **Temática**, v. 12, n. 01, p. 185–198, 2016.
- FAR CZADI, L. **Analysing your Strava data**. Disponível em: <<https://medium.com/epfl-extension-school/using-the-strava-api-and-python-d93d480490ef>>. Acesso em: 7 out. 2020.
- FIORDELLI, M.; DIVIANI, N.; SCHULZ, P. J. Mapping mHealth research: a decade of evolution. **J Med Internet Res**, v. 15, n. 5, p. e95, 2013.
- FLECK, M. P. A. et al. Aplicação da versão em português do instrumento abreviado de avaliação da qualidade de vida “WHOQOL-bref”. **Revista de Saúde Pública**, v. 34, n. 2, p. 178–183, 2000.
- FOX, K. R. et al. Physical activity and mental well-being in older people participating in the Better Ageing Project. **European journal of applied physiology**, v. 100, n. 5, p. 591–602, 2007.
- FREDERICSON, M.; MISRA, A. K. Epidemiology and aetiology of marathon running injuries. **Sports Medicine**, v. 37, n. 4–5, p. 437–439, 2007.
- GARBER, C. E. et al. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. **Medicine and Science in Sports and Exercise**, v. 43, n. 7, p. 1334–1359, 2011.
- GILES, D.; DRAPER, N.; NEIL, W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. **European Journal of Applied Physiology**, v. 116, n. 3, p. 563–571, 26 mar. 2016.
- GLADWELL, M. **Fora de série-Outliers**. Rio de Janeiro: Sextante, 2013.

GÖTSCHI, T.; GARRARD, J.; GILES-CORTI, B. Cycling as a Part of Daily Life: A Review of Health Perspectives. **Transport Reviews**, v. 36, n. 1, p. 45–71, 2 jan. 2016.

GUYON, I. et al. Gene selection for cancer classification using support vector machines. **Machine learning**, v. 46, n. 1–3, p. 389–422, 2002.

HAN, J.; KAMBER, M.; PEI, J. **Data mining concepts and techniques**. 3. ed. Waltham, Mass.: Morgan Kaufmann Publishers, 2012.

HEYWARD, V. H. **Avaliação física e prescrição de exercício: técnicas avançadas**. Porto Alegre: Artmed, 2013.

HINO, A. A. F. et al. Prevalência de lesões em corredores de rua e fatores associados. **Revista Brasileira de Medicina do Esporte**, v. 15, n. 1, p. 36–39, fev. 2009.

HÖPPE, P. The physiological equivalent temperature: a universal index for the biometeorological assessment of the thermal environment. **International Journal of Biometeorology**, 1999.

KOSHY, A. N. et al. Smart watches for heart rate assessment in atrial arrhythmias. **International journal of cardiology**, v. 266, p. 124–127, 2018.

KUHN, M.; JOHNSON, K. **Feature engineering and selection: A practical approach for predictive models**. Florida: CRC Press, 2019.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação gerenciais**. 11. ed. São Paulo: Pearson Education do Brasil, 2014.

LAVIE, C. J. et al. Effects of Running on Chronic Diseases and Cardiovascular and All-Cause Mortality. **Mayo Clinic Proceedings**, v. 90, n. 11, p. 1541–1552, nov. 2015.

LEE, D. et al. Running as a Key Lifestyle Medicine for Longevity. **Progress in Cardiovascular Diseases**, v. 60, n. 1, p. 45–55, jul. 2017.

LEVINE, D.; BERENSON, M.; STEPHAN, D. Estatística: Teoria e aplicações usando Microsoft Excel em Português. Rio de Janeiro: LTC, n. 7, 2014.

LIMA, J. L. O.; ALVARES, L. **Organização e representação da informação e do conhecimento**. São Paulo: B4 Editores, 2012.

MADAKAM, S.; RAMASWAMY, R.; TRIPATHI, S. Internet of Things (IoT): A Literature Review. **Journal of Computer and Communications**, v. 03, n. 05, p. 164–173, 2015.

MARSHALL, L. Facilitating knowledge management and knowledge sharing: new opportunities for information professionals. **Online**, v. 21, n. 5, 1997.

- MARTI, B.; ABELIN, T.; MINDER, C. Relationship of Training and Life-Style to 16-km Running Time of 4000 Joggers. **International journal of sports medicine**, v. 9, n. 02, p. 85–91, 1988.
- MATZARAKIS, A.; MAYER, H.; IZIOMON, M. G. Applications of a universal thermal index: Physiological equivalent temperature. **International Journal of Biometeorology**, 1999.
- MCINTOSH, A. S. Risk compensation, motivation, injuries, and biomechanics in competitive sport. **British journal of sports medicine**, v. 39, n. 1, p. 2–3, 2005.
- MICROSOFT CORPORATION (ED.). **Microsoft Press computer user's dictionary**. Redmond, Wash: Microsoft Press, 1998.
- MORGAN, W.; COSTILL, D. Selected psychological characteristics and health behaviors of aging marathon runners: a longitudinal study. **International journal of sports medicine**, v. 17, n. 04, p. 305–312, 1996.
- MUMMERY, K.; SCHOFIELD, G.; CAPERCHIONE, C. Physical activity: Physical activity dose-response effects on mental health status in older adults. **Australian and New Zealand Journal of Public Health**, v. 28, n. 2, p. 188–192, 2004.
- NASH, O. **Mining the Strava data**. Disponível em: <<http://olivernash.org/2014/05/25/mining-the-strava-data/>>. Acesso em: 7 out. 2020.
- NATIONAL INSTITUTES OF HEALTH. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. **NIH Publication**, 1998.
- OIKAWA, E. Dinâmicas relacionais contemporâneas: visibilidade, performances e interações nas redes sociais da internet. **Interações em rede**, 2013.
- OJA, P. et al. Health benefits of cycling: A systematic review. **Scandinavian Journal of Medicine and Science in Sports**, v. 21, n. 4, p. 496–509, 2011.
- OLIVEIRA, R. **Polar e Strava agora tem sincronização automática**. Disponível em: <<https://canaldebike.com.br/2015/12/17/polar-e-strava-agora-tem-sincronizacao-automatica>>. Acesso em: 7 maio. 2020.
- ORTEGA, J. Issues on the application of data mining. **Artificial Intelligence Review—An International Science and Engineering Journal**, v. 14, 2000.
- PALLOTTA, M.; HERDIES, D. L.; GONÇALVES, L. G. G. Estudo das condições de tempo e conforto térmico no desempenho esportivo aplicado à Maratona da cidade do Rio De Janeiro. **Revista Brasileira de Meteorologia**, v. 30, n. 2, p. 223–240, 2015.
- PEZZOLI, A. et al. Analysis of the thermal comfort in cycling athletes. **Procedia Engineering**, v. 34, p. 433–438, 2012.

POWERS, S. K.; HOWLEY, E. T. **Fisiologia do exercício: teoria e aplicação ao condicionamento e ao desempenho**. Barueri: Manole, 2000.

PRESSLER, A.; NIEBAUER, J. (EDS.). **Textbook of Sports and Exercise Cardiology**. Cham: Springer International Publishing, 2020.

PUCCI, G. et al. Quality of life and physical activity among adults: population-based study in Brazilian adults. **Quality of Life Research**, v. 21, n. 9, p. 1537–1543, 1 nov. 2012a.

PUCCI, G. C. M. F. et al. Associação entre atividade física e qualidade de vida em adultos. **Revista de Saúde Pública**, v. 46, n. 1, p. 166–179, fev. 2012b.

PUJARI, A. K. **Data mining techniques**. Hyderabad, India: Universities press, 2001.

PURE FITNESS. **Apple GymKit**. Disponível em:
https://www.pure-fitness.com/hongkong/page/view/apple_gymkit. Acesso em: 7 maio. 2020.

RIEHLE, D. **Framework design: A role modeling approach**. Zurich: ETH Zurich, 2000.

ROSA, B. N. DE; SILVEIRA, E. DE C. **Manual de uso: Programa de Identidade Visual e Sistema de Sinalização**. Disponível em:
https://www2.unesp.br/portal#/aci_ses/normas-e-padroes/identidade-visual-da-unesp. Acesso em: 28 jan. 2020.

RUIZ, F.; OLIVEIRA, P. R. DE. Ciclismo de velocidade: meios de treinamento de força. **Conexões**, v. 6, p. 22–34, 14 jul. 2008.

SCHULZ, S. Recreational cyclists: The relationship between low back pain and training characteristics. **Journal of Science and Medicine in Sport**, v. 12, p. S45, 2009.

SERRANO, K. J. et al. Mining Health App Data to Find More and Less Successful Weight Loss Subgroups. **J Med Internet Res**, v. 18, n. 6, p. e154, 2016.

SETHI, P.; SARANGI, S. R. Internet of Things: Architectures, Protocols, and Applications. **Journal of Electrical and Computer Engineering**, v. 2017, 2017.

SILVA, S. L. Gestão do conhecimento: uma revisão crítica orientada pela abordagem da criação do conhecimento. **Ciência da Informação**, v. 33, n. 2, p. 143–151, 2005.

STEWART, K. J. et al. Are fitness, activity, and fatness associated with health-related quality of life and mood in older persons? **Journal of Cardiopulmonary Rehabilitation and Prevention**, v. 23, n. 2, p. 115–121, 2003.

STRAVA. **Strava**. Disponível em: <https://www.strava.com>. Acesso em: 1 set. 2020a.

STRAVA. Strava Milestones: 50 Million Athletes and 3 Billion Activity Uploads. Disponível em: <<https://blog.strava.com/press/strava-milestones-50-million-athletes-and-3-billion-activity-uploads/>>. Acesso em: 7 maio. 2020b.

SUPING, Z. et al. Study of the relationships between weather conditions and the marathon race, and of meteorotropic effects on distance runners. **International Journal of Biometeorology**, 1992.

SVETNIK, V. et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. **Journal of chemical information and computer sciences**, v. 43, n. 6, p. 1947–1958, 2003.

TANAKA, H.; MONAHAN, K. D.; SEALS, D. R. Age-predicted maximal heart rate revisited. **Journal of the American College of Cardiology**, v. 37, n. 1, p. 153–156, 2001.

THURAISINGHAM, B. M. **Data mining : technologies, techniques, tools, and trends**. Boca Raton: CRC Press, 1999.

TUOMI, I. Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. **Journal of Management Information Systems**, v. 16, n. 3, p. 103–117, 1999.

VAGETTI, G. C. et al. The association between physical activity and quality of life domains among older women. **Journal of aging and physical activity**, v. 23, n. 4, p. 524–533, 2015.

VAN GENT, R. et al. Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. **British journal of sports medicine**, v. 41, n. 8, p. 469–480, 2007.

WARBURTON, D. E. R.; BREDIN, S. S. D. Health benefits of physical activity: A systematic review of current systematic reviews. **Current Opinion in Cardiology**, v. 32, n. 5, p. 541–556, 2017.

WARBURTON, D. E. R.; NICOL, C. W.; BREDIN, S. S. D. Health benefits of physical activity: the evidence. **Canadian Medic Association Journal**, 2006.

WEINBERG, R. S.; GOULD, D. **Fundamentos da psicologia do esporte e do exercício**. 6. ed. Porto Alegre: Artmed, 2017.

WENDEL-VOS, G. C. W. et al. Leisure time physical activity and health-related quality of life: cross-sectional and longitudinal associations. **Quality of Life research**, v. 13, n. 3, p. 667–677, 2004.

WEST, L. R. Strava: Challenge yourself to greater heights in physical activity/cycling and running. **British Journal of Sports Medicine**, v. 49, n. 15, p. 1024, 2015.

WILLIAMS, K. **Top Grossing Health and Fitness Apps Worldwide for February 2020**. Disponível em: <<https://sensortower.com/blog/top-grossing-health-and-fitness-apps-february-2020>>. Acesso em: 9 maio. 2020.

WILLIAMS, P. T. Relationship of distance run per week to coronary heart disease risk factors in 8283 male runners: the National Runners' Health Study. **Archives of Internal Medicine**, v. 157, n. 2, p. 191–198, 1997.

WORLD HEALTH ORGANIZATION. Measuring quality of life: The development of the World Health Organization Quality of Life Instrument (WHOQOL). **Psychology Medicine**, 1998.

YANFEN, L.; PU, H. Smart Sportswear. **2011 International Conference on Future Computer Science and Education**, p. 135–138, 2011.

YE, N. **The Handbook of Data Mining**. New Jersey: Human factors and ergonomics, 2003.

ZENG, L.; ALMQUIST, Z. W.; SPIRO, E. S. Let's Workout! Exploring Social Exercise in an Online Fitness Community. **iConference 2017**, p. 1–12, 2015.

APÊNDICE A – TCLE (português)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Conselho Nacional de Saúde, Resolução 466/2012 CNS/MS

Você está sendo convidado(a) a participar de uma pesquisa de Doutorado intitulada “MINERAÇÃO DE DADOS E DESCOBERTA DE CONHECIMENTO A PARTIR DE DADOS DE ATIVIDADE FÍSICA, SAÚDE, CLIMA E QUALIDADE DE VIDA”, que será desenvolvida por Rafael Macedo Sulino, RG nº 6.817.703, do Programa de Pós-graduação em Desenvolvimento Humano e Tecnologias da Unesp de Rio Claro, sob a orientação do Prof. Dr. Carlos Norberto Fischer, Professor Assistente Doutor do Departamento de Estatística, Matemática Aplicada e Computação (DEMAC), do Instituto de Geociências e Ciências Exatas (IGCE) da Universidade Estadual Paulista (Unesp), campus de Rio Claro/SP, e coorientada pelo Prof. Dr. Wellington Roberto Gomes Carvalho do Departamento de Educação Física da Universidade Federal de Uberlândia. Este documento, chamado Termo de Consentimento Livre e Esclarecido, visa assegurar seus direitos e deveres como participante.

A pesquisa tem como objetivo propor uma plataforma capaz de promover a descoberta de conhecimentos relevantes, por meio da análise de dados coletados por aplicativos e dispositivos de monitoramento de atividade física e saúde, além de dados climáticos. Esta plataforma poderá ser utilizada tanto por praticantes de atividade física quanto por profissionais que orientam a prática de atividade física, de forma a auxiliar na tomada de decisões para o planejamento de uma prática mais eficiente e segura.

Por essa razão, o(a) senhor(a) está sendo convidado(a) a participar da pesquisa. Caso o(a) senhor(a) aceite participar da pesquisa, sua participação consistirá em:

1. Preencher o questionário online denominado WHOQOL-BREF, um instrumento desenvolvido pela Organização Mundial da Saúde, que contém 26 perguntas sobre saúde e qualidade de vida. O tempo a ser gasto para responder ao questionário será cerca de 10 minutos.
2. Autorizar o compartilhamento de seus dados registrados em aplicativos ou dispositivos de monitoramento de atividade física e saúde.

SIGILO E PRIVACIDADE

Todas as informações que o(a) senhor(a) nos fornecer serão registradas em banco de dados, em formato digital, e serão utilizadas somente para esta pesquisa. Sua identidade será codificada por meio de chave numérica, de forma a impossibilitar sua identificação nominal. Seu endereço IP não será registrado, e seus dados estarão protegidos no banco de dados por meio de criptografia e firewall.

Suas respostas e dados pessoais serão guardados na confidencialidade, seu nome não aparecerá durante a pesquisa, nem quando os resultados forem apresentados, e seus dados de contato não serão repassados a outrem.

RISCOS DA PESQUISA

Responder ao questionário não está associado a riscos de saúde diretos. Os riscos existentes, mesmo que mínimos, são relacionados a invasão de privacidade e vazamento de informações confidenciais. Serão adotadas as seguintes medidas, com o objetivo de minimizar tais riscos: seus dados serão registrados em banco de dados, e as variáveis sensíveis serão criptografadas. Sua identificação será convertida em chave numérica, de forma a impossibilitar qualquer identificação nominal. Seu endereço IP não será registrado. Ao preencher e enviar seus dados, os mesmos estarão protegidos por meio de certificado SSL. O servidor onde está o banco de dados será protegido por firewall além de outras medidas de segurança oferecidas pelo datacenter.

A qualquer momento, antes, durante ou após a participação, coloco-me à disposição para esclarecimentos sobre eventuais dúvidas relacionadas à pesquisa. A participação nesta pesquisa é voluntária, e sua recusa em participar não lhe trará nenhum dano ou punição. Você poderá se recusar a participar, ou retirar seu consentimento, em qualquer fase da pesquisa, sem penalização alguma.

BENEFÍCIOS

Este estudo não apresenta benefícios diretos e imediatos ao participante. Porém, seus resultados poderão ser benéficos no futuro, no sentido de proporcionar conhecimentos relevantes, que possam ser utilizados tanto por praticantes de atividade física quanto profissionais que orientam a prática de atividade física, de forma a auxiliar na tomada de decisões para o planejamento de uma prática mais eficiente e segura.

CONSENTIMENTO

Se o(a) senhor(a) se sentir esclarecido(a) sobre a pesquisa, seus objetivos, eventuais riscos e benefícios, convido-o(a) a assinar eletronicamente este Termo (clicando no botão "Quero Participar"), e guardar uma via assinada eletronicamente pelo pesquisador responsável.

Rio Claro, 7 de dezembro de 2018.

DADOS SOBRE A PESQUISA

Título do Projeto: MINERAÇÃO DE DADOS E DESCOBERTA DE CONHECIMENTO A PARTIR DE DADOS DE ATIVIDADE FÍSICA, SAÚDE, CLIMA E QUALIDADE DE VIDA

Pesquisador responsável: Rafael Macedo Sulino

Cargo/função: Doutorando em Desenvolvimento Humano e Tecnologias

Instituição: UNESP – Campus de Rio Claro

Endereço: Avenida 24A, nº 1515, Instituto de Biociências – IB

Dados para contato: telefone (35) 99167-5689 / e-mail: rafael.sulino@unesp.br

CEP-IB/UNESP-CRC

Avenida 24A, nº 1515 – Bela Vista – CEP 13506-900 – Rio Claro/SP

Telefone: (19) 3526-9678

Número do parecer: **3.064.004**

APÊNDICE B – TCLE (inglês)

FREE AND INFORMED CONSENT TERM

Brazilian National Health Council, Resolution 466/2012

You are being invited to participate in a PhD research study entitled "DATA MINING AND DISCOVERY OF KNOWLEDGE FROM PHYSICAL ACTIVITY, HEALTH, CLIMATE AND QUALITY OF LIFE DATA", which will be developed by Rafael Macedo Sulino, ID document number 6817703, of the Graduate Program in Human Development and Technologies of São Paulo State University, under the guidance of Carlos Norberto Fischer, PhD., Assistant Professor, Department of Statistics, Applied Mathematics and Computing (DEMAC), Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (Unesp), campus of Rio Claro, and cooriented by Wellington Roberto Gomes Carvalho, PhD. of the Physical Education Department of the Federal University of Uberlândia, Minas Gerais. This document, called the Free and Informed Consent Term, aims to ensure your rights and duties as a participant.

The research aims to propose a platform capable of promoting the discovery of relevant knowledge, through the analysis of data collected by applications and devices for monitoring physical activity and health, as well as climatic data. This platform can be used by both physical activity practitioners and professionals who guide the practice of physical activity, in order to assist in making decisions for planning a more efficient and safe practice.

For this reason, you are being invited to participate in the survey. If you agree to participate in the research, your participation will consist of:

1. Complete the WHOQOL-BREF online questionnaire, an instrument developed by the World Health Organization, which contains 26 questions on health and quality of life. The time to be spent to respond to the questionnaire will be about 10 minutes.
2. Authorize the sharing of your recorded data in applications or devices for monitoring physical activity and health.

SIGIL AND PRIVACY

All information that you provide us will be recorded in a database, in digital format, and will be used only for this research. Your identity will be coded by a numeric key, in order to prevent its nominal identification. Your IP address will not be logged, and your data will be protected in the database through encryption and firewall.

Your responses and personal data will be kept confidential, your name will not appear during the survey, or when the results are presented, and your contact information will not be passed on to others.

RESEARCH RISKS

Responding to the questionnaire is not associated with direct health risks. Existing risks, even minimal ones, are related to invasion of privacy and leakage of confidential information. The following measures will be adopted in order to minimize such risks: your data will be recorded in a database, and sensitive variables will be encrypted. Its identification will be converted into a numerical key, in order to prevent any nominal identification. Your IP

address will not be logged. When completing and sending your data, they will be protected by an SSL certificate. The server on which the database is located will be protected by firewall as well as other security measures offered by the datacenter.

At any time, before, during or after the participation, I am available to clarify any questions related to the research. Participation in this research is voluntary, and your refusal to participate will not bring you any harm or punishment. You may refuse to participate, or withdraw your consent, at any stage of the research, without penalty.

BENEFITS

This study does not present direct and immediate benefits to the participant. However, its results may be beneficial in the future, in the sense of providing relevant knowledge that can be used by both physical activity practitioners and professionals that guide the practice of physical activity, in order to assist in decision making for planning a practice more efficiently and safely.

CONSENT

If you feel clear about the research, its objectives, possible risks and benefits, I invite you to electronically sign this Term (by clicking on the "I want to Participate" button), and save a copy electronically signed by the responsible researcher.

Rio Claro – São Paulo – Brazil, December 7, 2018.

RESEARCH DATA

Project Title: DATA MINING AND DISCOVERY OF KNOWLEDGE FROM PHYSICAL ACTIVITY,
HEALTH, CLIMATE AND QUALITY OF LIFE DATA

Researcher in charge: Rafael Macedo Sulino

Position / Function: Doctorate in Human Development and Technologies

Institution: São Paulo State University (UNESP) - Campus of Rio Claro

Address: Avenida 24A, n° 1515, Institute of Biosciences - IB

Contact information: phone +55 35 99167-5689 / e-mail: rafael.sulino@unesp.br

CEP-IB / UNESP-CRC

Avenida 24A, n° 1515 – Bela Vista – CEP 13506-900 – Rio Claro / SP

Phone: +55 19 3526-9678

Protocol number: **3.064.004**

APÊNDICE C – TCLE (espanhol)

TÉRMINO DE CONSENTIMIENTO LIBRE Y ESCLARECIDO (TCLE)

Consejo Nacional de Salud, Resolución 466/2012 CNS/MS (Brasil)

Usted está siendo invitado a participar en una investigación de Doctorado titulada "MINERÍA DE DATOS Y DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE DATOS DE ACTIVIDAD FÍSICA, SALUD, CLIMA Y CALIDAD DE VIDA", que será desarrollada por Rafael Macedo Sulino, RG nº 6.817.703, del Programa de Postgrado en Desarrollo Humano y Tecnologías de la Unesp de Río Claro, bajo la orientación del Prof. El Dr. Carlos Norberto Fischer, Profesor Asistente Doctor del Departamento de Estadística, Matemática Aplicada y Computación (DEMAC), del Instituto de Geociencias y Ciencias Exactas (IGCE) de la Universidad Estadual Paulista (Unesp), campus de Rio Claro / SP, y coorientada por el Instituto Prof. Dr. Wellington Roberto Gomes Carvalho del Departamento de Educación Física de la Universidad Federal de Uberlândia. Este documento, llamado Término de Consentimiento Libre y Esclarecido, pretende asegurar sus derechos y deberes como participante.

La investigación tiene como objetivo proponer una plataforma capaz de promover el descubrimiento de conocimientos relevantes, a través del análisis de datos recolectados por aplicaciones y dispositivos de monitoreo de actividad física y salud, además de datos climáticos. Esta plataforma podrá ser utilizada tanto por practicantes de actividad física como por profesionales que orientan la práctica de actividad física, para auxiliar en la toma de decisiones para la planificación de una práctica más eficiente y segura.

Por esta razón, usted está siendo invitado a participar en la investigación. Si usted aceptó participar en la investigación, su participación consistirá en:

1. Llenar el cuestionario online denominado WHOQOL-BREF, un instrumento desarrollado por la Organización Mundial de la Salud, que contiene 26 preguntas sobre salud y calidad de vida. El tiempo que se va a utilizar para responder al cuestionario será de unos 10 minutos.
2. Autorizar el uso compartido de sus datos registrados en aplicaciones o dispositivos de monitoreo de actividad física y salud.

SIGILO Y PRIVACIDAD

Toda la información que usted nos proporcionará será registrada en base de datos, en formato digital, y será utilizada solamente para esta investigación. Su identidad será codificada por medio de clave numérica, de forma a imposibilitar su identificación nominal. Su dirección IP no se registrará, y sus datos estarán protegidos en la base de datos a través de encriptación y firewall.

Sus respuestas y datos personales serán guardados en la confidencialidad, su nombre no aparecerá durante la investigación, ni cuando los resultados sean presentados, y sus datos de contacto no serán repasados a otro.

RIESGOS DE LA INVESTIGACIÓN

Responder al cuestionario no está asociado a riesgos de salud directos. Los riesgos existentes, aunque mínimos, están relacionados con la invasión de la privacidad y la filtración de información confidencial. Se adoptarán las siguientes medidas, con el objetivo de minimizar tales riesgos: sus datos se registrarán en una base de datos y se cifrarán las variables sensibles. Su identificación será convertida en clave numérica, de forma que imposibilite cualquier identificación nominal. Su dirección IP no se registrará. Al llenar y enviar sus datos, los mismos estarán protegidos por medio de un certificado SSL. El servidor donde está la base de datos será protegida por el firewall además de otras medidas de seguridad ofrecidas por el centro de datos.

En cualquier momento, antes, durante o después de la participación, me pongo a disposición para aclaraciones sobre eventuales dudas relacionadas con la investigación. La participación en esta investigación es voluntaria, y su rechazo a participar no le traerá ningún daño o castigo. Usted podrá negarse a participar, o retirar su consentimiento, en cualquier fase de la investigación, sin penalización alguna.

VENTAJAS

Este estudio no presenta beneficios directos e inmediatos al participante. Sin embargo, sus resultados podrán ser beneficiosos en el futuro, en el sentido de proporcionar conocimientos relevantes, que puedan ser utilizados tanto por practicantes de actividad física como profesionales que orientan la práctica de actividad física, de forma a auxiliar en la toma de decisiones para la planificación de una práctica más eficiente y segura.

CONSENTIMIENTO

Si tiene claro la investigación, sus objetivos, posibles riesgos y beneficios, lo invito a firmar electrónicamente este Término (haciendo clic en el botón “Quiero Participar”), y guardar copia firmada electrónicamente por el investigador responsable.

Rio Claro – São Paulo – Brasil, 7 de diciembre de 2018.

DATOS SOBRE LA INVESTIGACIÓN

Título del Proyecto: MINERÍA DE DATOS Y DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE DATOS DE ACTIVIDAD FÍSICA, SALUD, CLIMA Y CALIDAD DE VIDA

Investigador responsable: Rafael Macedo Sulino

Cargo / función: Doctorado en Desarrollo Humano y Tecnologías

Institución: UNESP - Campus de Rio Claro

Dirección: Avenida 24A, nº 1515, Instituto de Biociencias – IB

Datos de contacto: teléfono +55 35 99167-5689 / e-mail: rafael.sulino@unesp.br

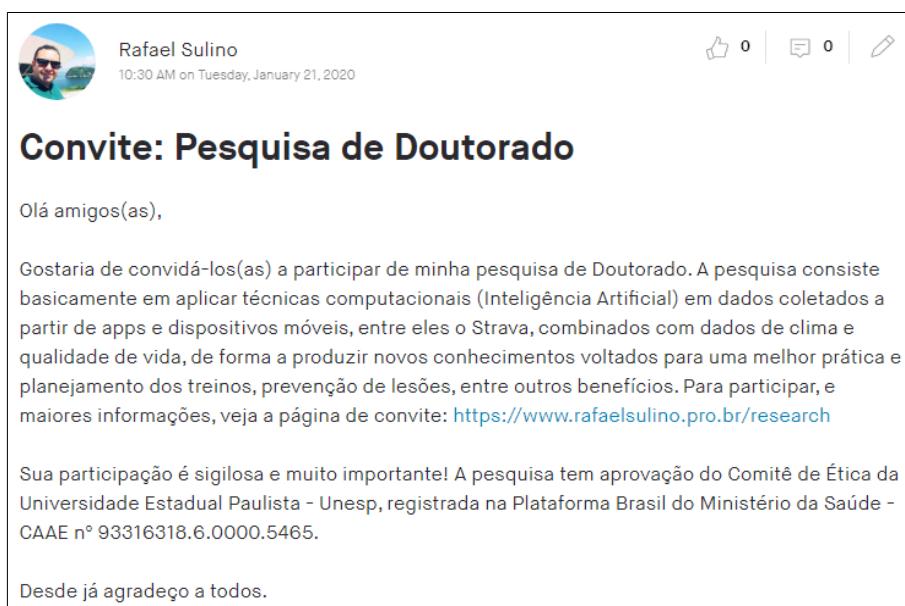
CEP-IB / UNESP-CRC

De la ciudad de Rio Claro, São Paulo, Brasil.

Teléfono: +55 19 3526-9678

Número del dictamen: **3.064.004**

APÊNDICE D – Convite rede social do STRAVA (português)



Rafael Sulino
10:30 AM on Tuesday, January 21, 2020

Convite: Pesquisa de Doutorado

Olá amigos(as),

Gostaria de convidá-los(as) a participar de minha pesquisa de Doutorado. A pesquisa consiste basicamente em aplicar técnicas computacionais (Inteligência Artificial) em dados coletados a partir de apps e dispositivos móveis, entre eles o Strava, combinados com dados de clima e qualidade de vida, de forma a produzir novos conhecimentos voltados para uma melhor prática e planejamento dos treinos, prevenção de lesões, entre outros benefícios. Para participar, e maiores informações, veja a página de convite: <https://www.rafaelsulino.pro.br/research>

Sua participação é sigilosa e muito importante! A pesquisa tem aprovação do Comitê de Ética da Universidade Estadual Paulista - Unesp, registrada na Plataforma Brasil do Ministério da Saúde - CAAE nº 93316318.6.0000.5465.

Desde já agradeço a todos.

APÊNDICE E – Convite rede social do STRAVA (inglês)



Rafael Sulino
10:07 AM on Monday, January 20, 2020

Invitation: PhD Research

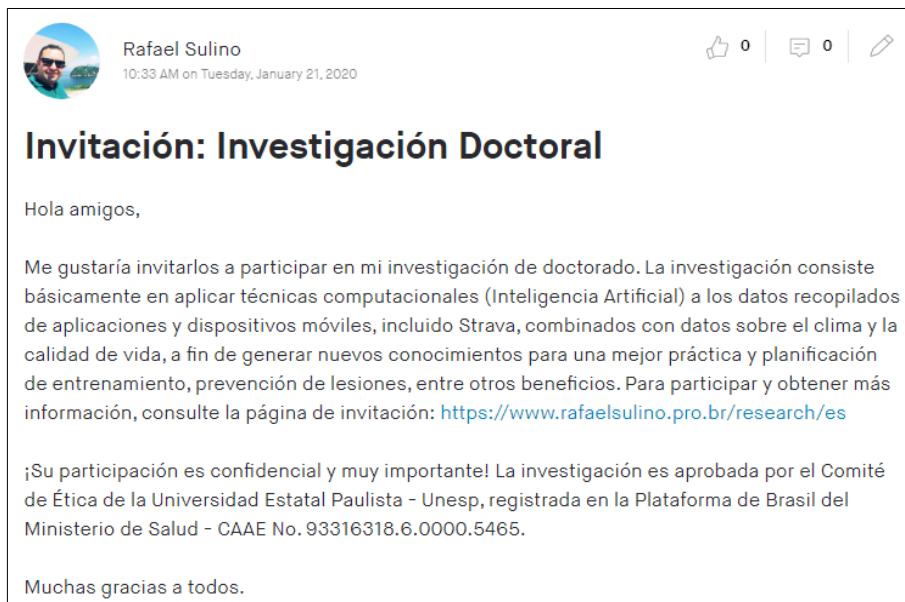
Hello friends,

I would like to invite you to take part in my PhD research. The research consists basically of applying computational techniques (Artificial Intelligence) to data collected from apps and mobile devices, among them Strava, in order to generate knowledge and information aimed at better practice and training planning, injury prevention, among other benefits. For more information, see the invitation webpage below: <https://www.rafaelsulino.pro.br/research/en>

Your participation is confidential and very important! The research is approved by the Ethics Committee of the São Paulo State University - Unesp, registered in the Brazil Platform of the Ministry of Health - CAAE No. 93316318.6.0000.5465.

I thank you all.

APÊNDICE F – Convite rede social do STRAVA (espanhol)



Rafael Sulino
10:33 AM on Tuesday, January 21, 2020

Invitación: Investigación Doctoral

Hola amigos,

Me gustaría invitarlos a participar en mi investigación de doctorado. La investigación consiste básicamente en aplicar técnicas computacionales (Inteligencia Artificial) a los datos recopilados de aplicaciones y dispositivos móviles, incluido Strava, combinados con datos sobre el clima y la calidad de vida, a fin de generar nuevos conocimientos para una mejor práctica y planificación de entrenamiento, prevención de lesiones, entre otros beneficios. Para participar y obtener más información, consulte la página de invitación: <https://www.rafaelsulino.pro.br/research/es>

¡Su participación es confidencial y muy importante! La investigación es aprobada por el Comité de Ética de la Universidad Estatal Paulista - Unesp, registrada en la Plataforma de Brasil del Ministerio de Salud - CAAE No. 93316318.6.0000.5465.

Muchas gracias a todos.

APÊNDICE G – Convite rede social *Instagram* (português)

 **Rafael Sulino** @rmsulino
2 hours ago

Convite: Pesquisa de Doutorado
Gostaria de convidá-l@s a participar de minha pesquisa de Doutorado. A pesquisa consiste basicamente em aplicar técnicas computacionais (Inteligência Artificial) em dados do Strava, combinados com dados de clima e qualidade de vida, de forma a produzir novos conhecimentos voltados para uma melhor prática e planejamento dos treinos, prevenção de lesões, entre outros benefícios. Para participar, e maiores informações, veja a página de convite:
www.rafaelsulino.pro.br/research (link na Bio)
A pesquisa tem aprovação do Comitê de Ética da Universidade Estadual Paulista - Unesp, registrada na Plataforma Brasil do Ministério da Saúde - CAAE nº 93316318.6.0000.5465.
#stravabrasil #mtbbrasil #ciclismo #ciclismobrasil
#praquepedala #corridaderua #bikebr #bikebrasil #pedal
#borapedalar #aquipedala #mtbbr #biketrilha #brasilride
#meupedal #bikenafoto #clickopedal

[Read Less](#)

Convite: Pesquisa de Doutorado



APÊNDICE H – Convite rede social *Instagram* (inglês)

 **Rafael Sulino** @rmsulino
8 hours ago

Invitation: PhD Research

I would like to invite you to take part in my PhD research. The research consists basically of applying computational techniques (Artificial Intelligence) to data collected from apps and mobile devices, among them Strava, in order to generate knowledge and information aimed at better practice and training planning, injury prevention, among other benefits. For more information, see the invitation webpage below:
www.rafaelsulino.pro.br/research/en (link in Bio)
#strava #stravarun #stravarunning #stravaclub #stravabike
#stravacycling #stravaphoto #mtb #mountainbike #cycling
#running

[Read Less](#)

Invitation: PhD Research



APÊNDICE I – Especificações técnicas da plataforma *Fitness Tools*

A plataforma *Fitness Tools* conta com segurança no tráfego de dados por meio da tecnologia *Secure Socket Layer (SSL)*, um padrão global de segurança que cria um canal criptografado entre o servidor *web* e o *software* navegador de Internet do usuário, proporcionando sigilo e segurança na transmissão dos dados. Conta, também, com monitoramento constante da disponibilidade, por meio da ferramenta *Uptime Robot*, e registro de eventos relacionados a erros na aplicação (*log*).

A tela de gerenciamento de atletas (Figura I-1) exibe todos os atletas cadastrados (10 por página) e possibilita tanto o cadastro de um novo atleta como a atualização de dados de atletas já cadastrados.

Figura I-1 – Gerenciamento de dados dos atletas cadastrados.

Atletas		
Cadastrar novo		
« 1 2 3 4 5 ... 109 110 »		
Nome	Atividades	Ações
Atleta_4656551	▲	📅 📝 📋 ⏲️ ☀️
Atleta_16032303	▲	📅 📝 📋 ⏲️ ☀️
Atleta_18032312	▲	📅 📝 📋 ⏲️ ☀️
Atleta_17817467	▲	📅 📝 📋 ⏲️ ☀️
Atleta_10549309	▲	📅 📝 📋 ⏲️ ☀️
Atleta_22282074	▲	📅 📝 📋 ⏲️ ☀️
Atleta_5843603	▲	📅 📝 📋 ⏲️ ☀️
Atleta_8566249	▲	📅 📝 📋 ⏲️ ☀️
Atleta_20256320	▲	📅 📝 📋 ⏲️ ☀️
Atleta_7320730	▲	📅 📝 📋 ⏲️ ☀️

Fonte: Elaborado pelo autor

Cadastro de usuários da plataforma

Para ter acesso à plataforma *Fitness Tools*, o futuro usuário deve solicitar, por e-mail, o seu cadastro ao administrador do sistema. Após o cadastro de um usuário,

o requerente receberá, por e-mail, os dados de acesso, contendo uma senha gerada de forma automática, a qual deverá ser trocada obrigatoriamente no primeiro acesso. A partir daí, o novo usuário passa a ter acesso aos recursos da plataforma e poderá cadastrar seus clientes (atletas), obter os dados das atividades realizadas e realizar as análises.

Cadastro de atletas

O cadastro inicial de atletas é realizado por meio da aba “Cadastro” (Figura I-2) e consiste em inserir os dados básicos no banco do sistema, tais como nome, sexo, data de nascimento, massa corporal e estatura.

Figura I-2 – Cadastro de atletas

Dados cadastrais

Nome:
Rafael Sulino

Sexo:
 Masculino Feminino

Data de nascimento:
05/11/1979

Massa corporal em kg:
98

Estatura em centímetros:
172

Gravar

Fonte: Elaborado pelo autor

Ao gravar os dados, é realizada uma validação das informações inseridas (por exemplo, verifica se a data de nascimento é uma data válida). As informações referentes à anamnese são inseridas por meio da aba “Anamnese” (Figura I-3) e as referentes à qualidade de vida por meio da aba “Qualidade de vida” (Figura I-4).

Figura I-3 – Registro de dados referentes à anamnese

Cadastro	Anamnese	Qualidade de vida	Ocorrências	Conexões
----------	----------	-------------------	-------------	----------

Anamnese

1. Possui diagnóstico de Diabetes?

 Não

 Sim, mas não sei o tipo

 Sim, tipo I

 Sim, tipo II
2. Possui diagnóstico de Hipertensão Arterial?

 Sim Não
3. Possui diagnóstico de algum problema cardíaco?

 Sim Não
4. Possui histórico de doença cardiovascular em sua família?

 Sim Não
5. Utiliza algum medicamento de uso contínuo?

 Não

 Sim, para controle da pressão arterial

 Sim, para problema cardíaco

 Sim, para diabetes (via oral)

 Sim, para diabetes (insulina)

 Sim, outros

Fonte: Elaborado pelo autor

Figura I-4 – Registro de dados sobre qualidade de vida

Cadastro	Anamnese	Qualidade de vida	Ocorrências	Conexões
----------	----------	-------------------	-------------	----------

Qualidade de vida

[Imprimir o questionário WHOQOL-Bref](#)

1. Como você avaliaria sua qualidade de vida?

 Muito ruim

 Ruim

 Nem ruim nem boa

 Boa

 Muito boa
2. Quão satisfeito(a) você está com a sua saúde?

 Muito insatisfeito

 Insatisfeito

 Nem satisfeito nem insatisfeito

 Satisfeito

 Muito satisfeito
3. Em que medida você acha que sua dor (física) impede você de fazer o que você precisa?

 Nada

 Muito pouco

 Mais ou menos

 Bastante

 Extremamente

Fonte: Elaborado pelo autor

O preenchimento do formulário pode ser feito pelo profissional usuário da plataforma *Fitness Tools*, por meio da transcrição do questionário preenchido previamente na forma física pelo atleta ou pode ser feito pelo próprio atleta, por meio de um link gerado pelo profissional e compartilhado, de modo a possibilitar o acesso do atleta ao formulário eletrônico.

Registro de ocorrências

O registro de ocorrências é realizado por meio da aba “Ocorrências” (Figura I-5) e consiste no registro de fatos relevantes relacionados ao treinamento, tais como lesões, percepção de desconforto, entre outros.

Figura I-5 – Registro de ocorrências

The screenshot shows a user interface for managing occurrences. At the top, there is a horizontal navigation bar with five tabs: 'Cadastro', 'Anamnese', 'Qualidade de vida', 'Ocorrências' (which is highlighted in blue), and 'Conexões'. Below the navigation bar, the main content area is titled 'Ocorrências'. It contains the following fields:

- Data da ocorrência:** A date input field showing '05/09/2020' with a calendar icon to its right.
- Tipo de ocorrência:** A dropdown menu currently set to 'Lesão'.
- Descrição da ocorrência:** A text area containing the text: 'Possível lesão muscular. Durante o treino de ciclismo nesta data, o atleta relata ter sentido uma "fisgada". O treino foi interrompido para devida identificação e tratamento da lesão.'

At the bottom left of the form is a 'Gravar' (Save) button. A large red arrow points upwards from the bottom of the 'Ocorrências' tab towards the top of the page, indicating the active tab.

Fonte: Elaborado pelo autor

O registro de dados referentes às atividades realizadas (treinos) é feito usando a aba “Conexões” (Figura I-6), de forma a habilitar a obtenção de tais dados diretamente da API do *Strava*.

Figura I-6 – Integração com Strava e importação de arquivos

Fonte: Elaborado pelo autor

A inserção de novas atividades, registradas com o Strava, é realizada utilizando o botão “Atualizar” (Figura I-7), na tela de gerenciamento de atletas.

Figura I-7 – Atualização de dados de atividade física

Atletas

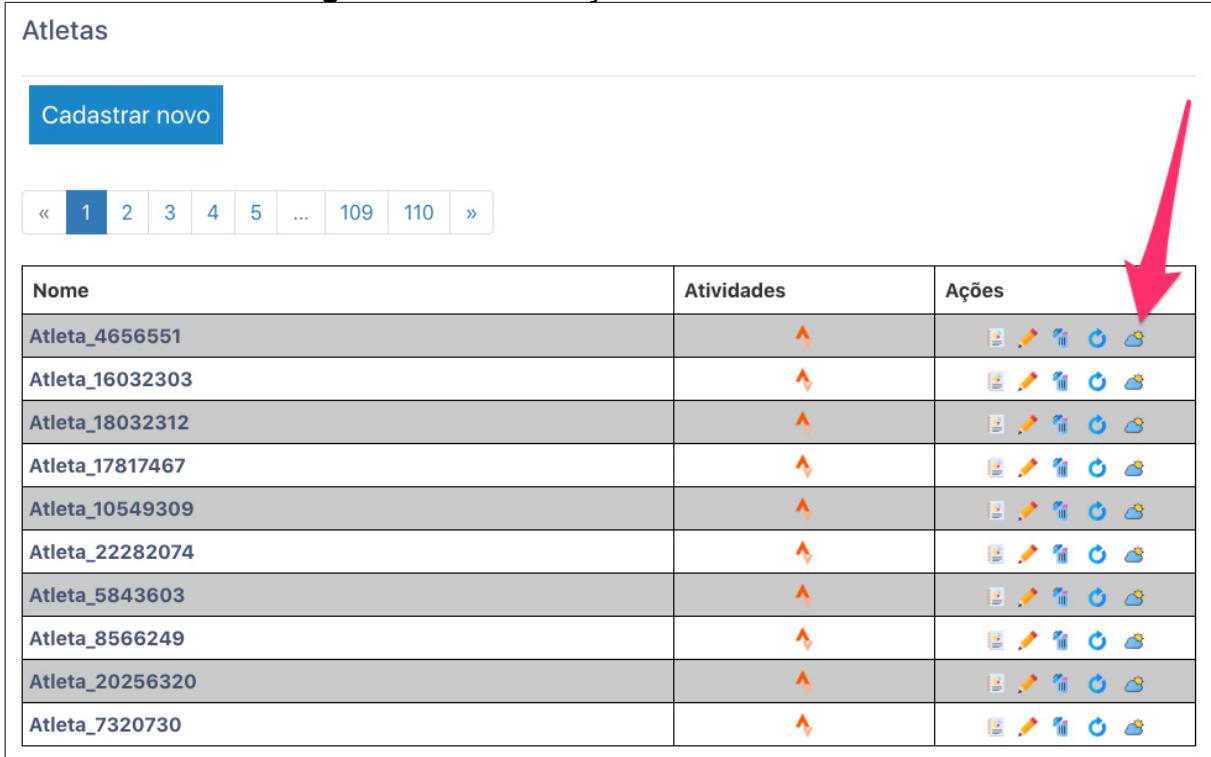
Cadastrar novo

Nome	Atividades	Ações
Atleta_4656551	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_16032303	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_18032312	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_17817467	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_10549309	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_22282074	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_5843603	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_8566249	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_20256320	▲	🕒 🖊️ 📈 ⚡ ☁
Atleta_7320730	▲	🕒 🖊️ 📈 ⚡ ☁

Fonte: Elaborado pelo autor

Os dados climáticos são integrados à plataforma por meio da API do serviço *World Weather Online*, sendo responsável pelo registro de dados climáticos referentes às atividades registradas, de forma complementar aos dados do *Strava*. O registro de dados climáticos é realizado por meio do botão “Atualizar dados climáticos” (Figura I-8) na tela de gerenciamento de atletas.

Figura I-8 – Atualização de dados climáticos



Nome	Atividades	Ações
Atleta_4656551	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_16032303	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_18032312	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_17817467	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_10549309	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_22282074	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_5843603	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_8566249	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_20256320	▲	🕒 📝 🏃‍♂️ 🌡️
Atleta_7320730	▲	🕒 📝 🏃‍♂️ 🌡️

Fonte: Elaborado pelo autor

As atividades registradas na plataforma *Fitness Tools* podem ser analisadas individualmente por meio do botão “Detalhes da atividade” (Figura I-9), na tela de gerenciamento das atividades.

Figura I-9 – Análise individual das atividades

Atividades

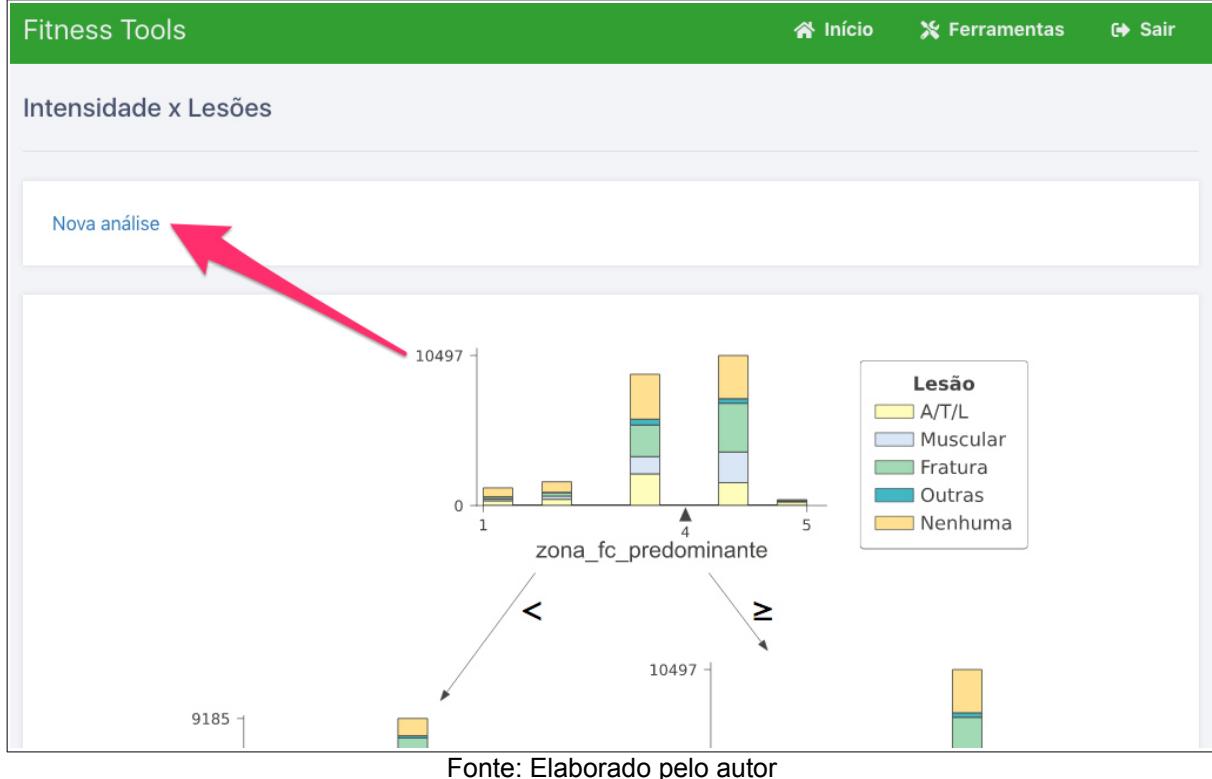
Data	Tipo	Ações
12/05/2019	Ride	
21/04/2019	Ride	
19/04/2019	Ride	
19/04/2019	Ride	
02/06/2018	Ride	
29/04/2018	Ride	
22/04/2018	Ride	
08/04/2018	Ride	
29/03/2018	Ride	
04/02/2018	Ride	



Fonte: Elaborado pelo autor

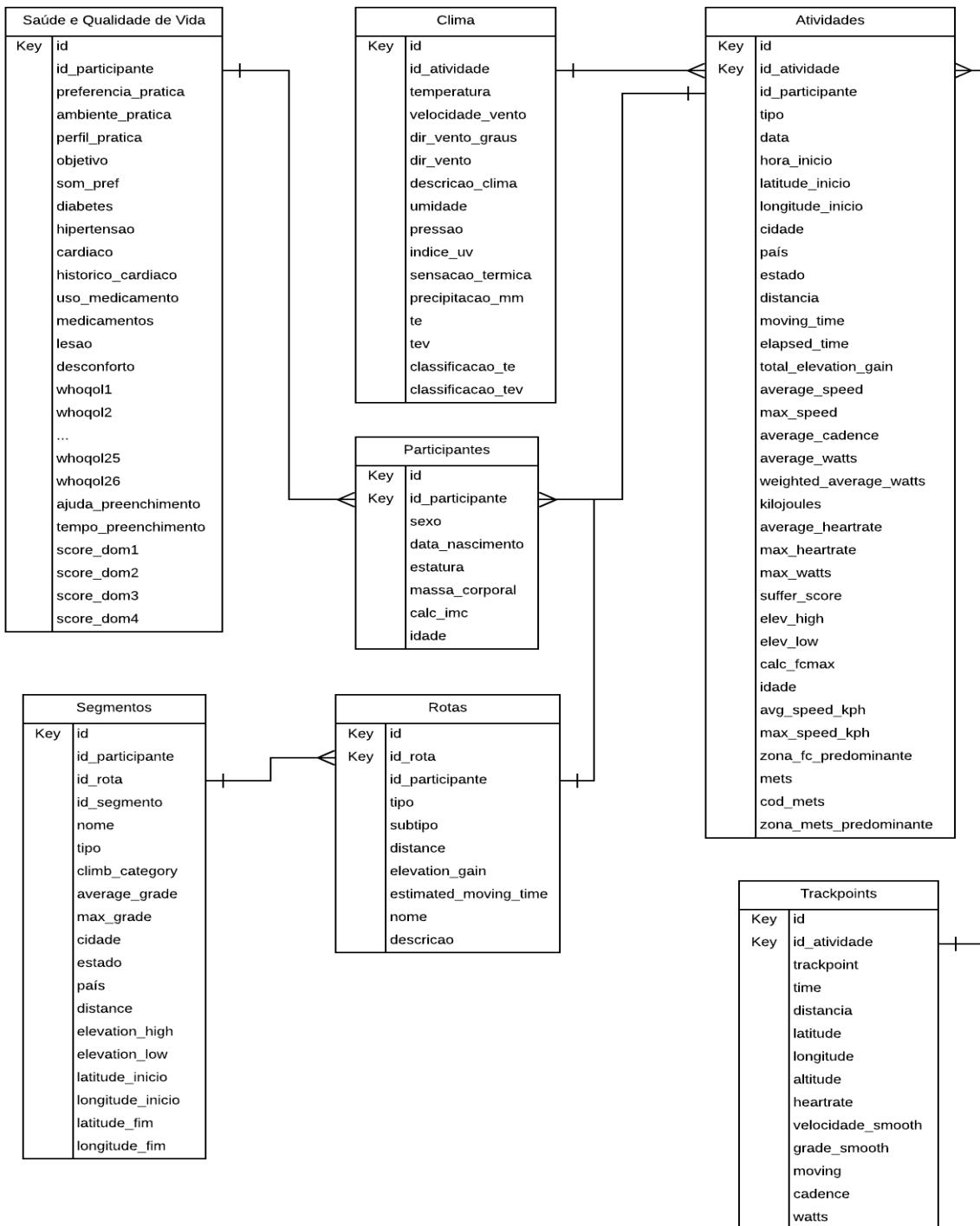
Mineração de dados

Após realizar uma análise por meio da ferramenta de mineração de dados, o usuário poderá facilmente iniciar outra análise usando o botão “Nova análise”, conforme destacado na Figura I-10.

Figura I-10 – Botão para nova análise (Mineração de Dados)

APÊNDICE J – Modelo Entidade Relacionamento da plataforma *Fitness Tools*

Modelo Entidade Relacionamento



APÊNDICE H – Questionário de Anamnese

Sobre você

1. Gênero: Masculino Feminino

2. Data de nascimento

dd / mm / aaaa



3. Massa corporal (peso) em kg

kg

4. Estatura em centímetros

cm

Sobre sua saúde

1. Possui diagnóstico de Diabetes?

- Não
- Sim, mas não sei o tipo
- Sim, tipo I
- Sim, tipo II

2. Possui diagnóstico de Hipertensão Arterial?

- Sim
- Não

3. Possui diagnóstico de algum problema cardíaco?

- Sim
- Não

4. Possui histórico de doença cardiovascular em sua família?

- Sim
- Não

5. Utiliza algum medicamento de uso contínuo?

- Não
- Sim, para controle da pressão arterial
- Sim, para problema cardíaco
- Sim, para diabetes (via oral)
- Sim, para diabetes (insulina)
- Sim, outros

6. Já sofreu algum tipo de lesão em decorrência da prática de atividade física?

- Não
- Sim, lesão muscular
- Sim, fratura
- Sim, articulação, tendão ou ligamento
- Sim, outras

7. Já sentiu algum tipo de desconforto durante a prática de atividade física?

- Não
- Sim, dor no peito
- Sim, tontura
- Sim, dor nas costas
- Sim, outros

Sobre sua atividade física

1. Qual sua preferência para a prática de atividade física?

- Individual
- Em grupo

2. Em que ambiente prefere praticar atividade física?

- Ao ar livre
- Ambiente fechado (academias ou clubes)

3. Qual o seu perfil como praticante de atividade física?

- Recreativo
- Atleta amador
- Atleta profissional

4. Qual o seu principal objetivo com a prática de atividade física?

- Saúde
- Socialização
- Desempenho / performance
- Competitivo

5. O que gosta de ouvir durante a prática de atividade física?

- Nada
- Música
- Podcast
- Audiolivro
- Outros

ANEXO A – Parecer consubstanciado do CEP

UNESP - INSTITUTO DE
BIOCIÊNCIAS DE RIO CLARO
DA UNIVERSIDADE ESTADUAL



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: MINERAÇÃO DE DADOS E DESCOBERTA DE CONHECIMENTO A PARTIR DE DADOS DE ATIVIDADE FÍSICA, SAÚDE E CLIMA

Pesquisador: RAFAEL MACEDO SULINO

Área Temática:

Versão: 2

CAAE: 93316318.6.0000.5465

Instituição Proponente: UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.064.004

Apresentação do Projeto:

Trata-se de uma pesquisa em nível de doutorado, que tem como temática a mineração de dados ligados à prática de atividade física, a ser desenvolvida pelo aluno RAFAEL MACEDO SULINO, sob a orientação do Prof. Dr. Carlos Norberto Fischer.

Objetivo da Pesquisa:

O objetivo geral deste estudo é propor uma plataforma capaz de promover a descoberta de conhecimentos relevantes, por meio da análise de dados coletados por aplicativos e dispositivos de monitoramento de atividade física e saúde, além de dados climáticos obtidos por meio de requisição em APIs (interfaces de programação) de acesso público.

Avaliação dos Riscos e Benefícios:

A pesquisa não oferece riscos diretos à saúde, já que se trata apenas da análise de dados produzidos pelo participante em sua rotina de prática de atividade física, sem qualquer intervenção. Devido à utilização do questionário WHOQOL-BREF, sobre saúde e qualidade de vida, os riscos, mesmo que mínimos, são relacionados a invasão de privacidade e vazamento de informações confidenciais do participante. Tais riscos serão tratados por meio da codificação da identidade do participante em chave numérica, bem como medidas de segurança e proteção dos dados, tais como criptografia, certificado SSL e firewall.

A pesquisa trará como benefício novos conhecimentos que possam auxiliar o profissional em uma

Endereço: Av.24-A n.º 1515	CEP: 13.506-900
Bairro: Bela Vista	
UF: SP	Município: RIO CLARO
Telefone: (19)3526-9678	Fax: (19)3534-0009
	E-mail: cepib@rc.unesp.br

**UNESP - INSTITUTO DE
BIOCIÊNCIAS DE RIO CLARO
DA UNIVERSIDADE ESTADUAL**



Continuação do Parecer: 3.064.004

melhor tomada de decisões quanto ao planejamento e prescrição de atividade física, de forma a torná-la mais segura e eficiente para os praticantes de atividade física.

Comentários e Considerações sobre a Pesquisa:

A pesquisa apresenta temática relevante e promissora para a contribuição de conhecimentos em relação às esferas da análise de dados e prática da atividade física. Está fundamentado em referências sobre mineração de dados e sua relação com o domínio da atividade física.

Considerações sobre os Termos de apresentação obrigatória:

Os termos apresentados estão em consonância com as orientações do CEP/UNESP-IB.

Conclusões ou Pendências e Lista de Inadequações:

O CEP REFERENDA O PARECER DO RELATOR:

"Considerando as correções sugeridas e apresentadas pelo pesquisador, considero o projeto aprovado".

Considerações Finais a critério do CEP:

O projeto encontra-se APROVADO para execução. Pedimos atenção aos seguintes itens:

- 1) De acordo com a Resolução CNS nº 466/12, o pesquisador deverá apresentar relatório final.
- 2) Eventuais emendas (modificações) ao protocolo devem ser apresentadas, com justificativa, ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada.
- 3) Sobre o TCLE: caso o termo tenha DUAS páginas ou mais, lembramos que no momento da sua assinatura, tanto o participante da pesquisa (ou seu representante legal) quanto o pesquisador responsável deverão RUBRICAR todas as folhas , colocando as assinaturas na última página.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJECTO_1097446.pdf	06/11/2018 07:42:43		Aceito
Outros	Correcoes.pdf	06/11/2018 07:42:23	RAFAEL MACEDO SULINO	Aceito
Projeto Detalhado / Brochura	Projeto_corrigido_20181106.pdf	06/11/2018 07:31:31	RAFAEL MACEDO SULINO	Aceito

Endereço: Av.24-A n.º 1515

Bairro: Bela Vista

CEP: 13.506-900

UF: SP

Município: RIO CLARO

Telefone: (19)3526-9678

Fax: (19)3534-0009

E-mail: cepib@rc.unesp.br

**UNESP - INSTITUTO DE
BIOCIÊNCIAS DE RIO CLARO
DA UNIVERSIDADE ESTADUAL**



Continuação do Parecer: 3.064.004

Investigador	Projeto_corrigido_20181106.pdf	06/11/2018 07:31:31	RAFAEL MACEDO SULINO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_20181106.pdf	06/11/2018 07:31:18	RAFAEL MACEDO SULINO	Aceito
Folha de Rosto	FolhaDeRosto_assinada.pdf	04/07/2018 20:13:20	RAFAEL MACEDO SULINO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

RIO CLARO, 07 de Dezembro de 2018

Assinado por:
Flávio Soares Alves
(Coordenador(a))

Endereço: Av.24-A n.º 1515	CEP: 13.506-900
Bairro: Bela Vista	
UF: SP	Município: RIO CLARO
Telefone: (19)3526-9678	Fax: (19)3534-0009
	E-mail: cepib@rc.unesp.br

ANEXO B – Decisão CEP/IB 01/2019



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Biociências – Câmpus de Rio Claro
Seção Técnica Acadêmica



COMITÉ DE ÉTICA EM PESQUISA
CEP-IB-UNESP- RIO CLARO

DIRETRIZES PARA A PESQUISA ON-LINE

DECISÃO CEP 01/19 - APROVADA NA 130ª REUNIÃO ORDINÁRIA, REALIZADA EM 12/12/19.

1. A abordagem e a aplicação da pesquisa devem ser feitas, preferencialmente, de modo presencial e pelo pesquisador responsável pela pesquisa.
2. Quando **não houver** a possibilidade de se operar a abordagem presencial, o CEP analisará a proposta on-line, mediante apresentação de justificativa pertinente, na submissão do protocolo de pesquisa, explicando os motivos pela opção on-line.
3. Só serão permitidas pesquisas on-line que envolvam riscos mínimos, **passíveis de serem controlados à distância**.
4. Como regra geral, não serão permitidas pesquisas exclusivamente online com menores de 18 anos ou outros grupos vulneráveis.
5. No item “Metodologia”, do projeto de pesquisa e das IBPs, o pesquisador deverá detalhar **todos os procedimentos on-line que serão utilizados**, tanto na seleção dos participantes, abordagem dos mesmos, quanto na aplicação da pesquisa.
6. Nos casos em que a abordagem on-line for feita por e-mail, anexar ao protocolo de pesquisa, registrado na Plataforma Brasil, o texto padrão que será enviado no e-mail endereçado aos participantes convidados da pesquisa.
7. Os TCLEs que exigirem datação e assinatura física formal, registrada em papel, poderão ser enviados por e-mail, pelo pesquisador responsável, desde que:
 - 7.1. Estejam digitalizados e já contendo a assinatura do pesquisador.
 - 7.2. O participante convidado devolva o termo assinado, datado e digitalizado, pelo mesmo e-mail no qual recebeu o convite para participação na pesquisa.
8. Poderá ser permitida a substituição do TCLE formal digitalizado e assinado, pelo TCLE virtual, desde que o sistema eletrônico utilizado permita:
 - 8.1. Apresentação de todas as informações relativas à pesquisa, em linguagem clara, acessível e em formato de convite (dados da pesquisa e do pesquisador, contatos, objetivos, benefícios, riscos, procedimentos de minimização, termos da participação voluntária).

Instituto de Biociências – Seção Técnica Acadêmica
Avenida 24-A nº 1515 - CEP 13506-900 - Rio Claro - S.P. - Brasil
☎ (19) 3526-9678 E cepib@rc.unesp.br



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Instituto de Biociências – Câmpus de Rio Claro
Seção Técnica Acadêmica



8.2. Apresentação de campo onde o participante possa manifestar seu aceite ou não em participar da pesquisa, do tipo:

- "Aceito () Não aceito ()".
- "*Ao clicar no botão abaixo, o(a) Senhor(a) concorda em participar da pesquisa nos termos deste TCLE. Caso não concorde em participar, apenas feche essa página no seu navegador*" ().

9. Além das informações habituais que devem constam no TCLE formal, o TCLE virtual deve informar que:

9.1. O participante receberá uma cópia do termo de consentimento com a datação e a assinatura física formal do pesquisador responsável. A cópia poderá ser impressa, ou digitalizada e pode ser fornecida em um dispositivo de armazenamento eletrônico, ou via e-mail.

9.2. Os dados coletados somente serão utilizados (para fins científicos) após o recebimento do TCLE assinado e datado pelo participante da pesquisa.

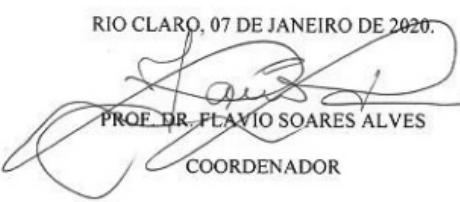
9.3. Caso o participante solicite, sua resposta deverá ser retirada da pesquisa, sem comprometimento do anonimato.

10. Quanto ao formulário/instrumento de pesquisa enviado on-line para o participante convidado:

10.1. Deverá ser totalmente anônimo, portanto, não poderá coletar nenhuma informação de identidade.

11. Casos não previstos por esta diretriz serão analisados pelo CEP.

RIO CLARO, 07 DE JANEIRO DE 2020.


PROF. DR. FLÁVIO SOARES ALVES

COORDENADOR

ANEXO C – Questionário WHOQOL Abreviado

WHOQOL - ABREVIADO

Versão em Português

PROGRAMA DE SAÚDE MENTAL
ORGANIZAÇÃO MUNDIAL DA SAÚDE
GENEBRA

Coordenação do GRUPO WHOQOL no Brasil

Dr. Marcelo Pio de Almeida Fleck
Professor Adjunto
Departamento de Psiquiatria e Medicina Legal
Universidade Federal do Rio Grande do Sul
Porto Alegre – RS - Brasil

Instruções

Este questionário é sobre como você se sente a respeito de sua qualidade de vida, saúde e outras áreas de sua vida. **Por favor, responda a todas as questões**. Se você não tem certeza sobre que resposta dar em uma questão, por favor, escolha entre as alternativas a que lhe parece mais apropriada. Esta, muitas vezes, poderá ser sua primeira escolha.

Por favor, tenha em mente seus valores, aspirações, prazeres e preocupações. Nós estamos perguntando o que você acha de sua vida, tomando como referência as **duas últimas semanas**. Por exemplo, pensando nas últimas duas semanas, uma questão poderia ser:

	nada	muito pouco	médio	muito	completamente
Você recebe dos outros o apoio de que necessita?	1	2	3	4	5

Você deve circular o número que melhor corresponde ao quanto você recebe dos outros o apoio de que necessita nestas últimas duas semanas. Portanto, você deve circular o número 4 se você recebeu "muito" apoio como abaixo.

	nada	muito pouco	médio	muito	completamente
Você recebe dos outros o apoio de que necessita?	1	2	3	4	5

Você deve circular o número 1 se você não recebeu "nada" de apoio.

Por favor, leia cada questão, veja o que você acha e circule no número e lhe parece a melhor resposta.

		muito ruim	ruim	nem ruim nem boa	boa	muito boa
1	Como você avaliaria sua qualidade de vida?	1	2	3	4	5

		muito insatisfeito	insatisfeito	nem satisfeito nem insatisfeito	satisfeito	muito satisfeito
2	Quão satisfeita(a) você está com a sua saúde?	1	2	3	4	5

As questões seguintes são sobre **o quanto** você tem sentido algumas coisas nas últimas duas semanas.

		nada	muito pouco	mais ou menos	bastante	extremamente
3	Em que medida você acha que sua dor (física) impede você de fazer o que você precisa?	1	2	3	4	5
4	O quanto você precisa de algum tratamento médico para levar sua vida diária?	1	2	3	4	5
5	O quanto você aproveita a vida?	1	2	3	4	5
6	Em que medida você acha que a sua vida tem sentido?	1	2	3	4	5
7	O quanto você consegue se concentrar?	1	2	3	4	5
8	Quão seguro(a) você se sente em sua vida diária?	1	2	3	4	5
9	Quão saudável é o seu ambiente físico (clima, barulho, poluição, atrativos)?	1	2	3	4	5

As questões seguintes perguntam sobre **quão completamente** você tem sentido ou é capaz de fazer certas coisas nestas últimas duas semanas.

		nada	muito pouco	médio	muito	completamente
10	Você tem energia suficiente para seu dia-a-dia?	1	2	3	4	5
11	Você é capaz de aceitar sua aparência física?	1	2	3	4	5
12	Você tem dinheiro suficiente para satisfazer suas necessidades?	1	2	3	4	5
13	Quão disponíveis para você estão as informações que precisa no seu dia-a-dia?	1	2	3	4	5
14	Em que medida você tem oportunidades de atividade de lazer?	1	2	3	4	5

As questões seguintes perguntam sobre **quão bem ou satisfeito** você se sentiu a respeito de vários aspectos de sua vida nas últimas duas semanas.

		muito ruim	ruim	nem ruim nem bom	bom	muito bom
15	Quão bem você é capaz de se locomover?	1	2	3	4	5

		muito insatisfeito	insatisfeito	nem satisfeito nem insatisfeito	satisfeito	muito satisfeito
16	Quão satisfeito(a) você está com o seu sono?	1	2	3	4	5
17	Quão satisfeito(a) você está com sua capacidade de desempenhar as atividades do seu dia-a-dia?	1	2	3	4	5
18	Quão satisfeito(a) você está com sua capacidade para o trabalho?	1	2	3	4	5
19	Quão satisfeito(a) você está consigo mesmo?	1	2	3	4	5
20	Quão satisfeito(a) você está com suas relações pessoais (amigos, parentes, conhecidos, colegas)?	1	2	3	4	5
21	Quão satisfeito(a) você está com sua vida sexual?	1	2	3	4	5
22	Quão satisfeito(a) você está com o apoio que você recebe de seus amigos?	1	2	3	4	5
23	Quão satisfeito(a) você está com as condições do local onde mora?	1	2	3	4	5
24	Quão satisfeito(a) você está com o seu acesso aos serviços de saúde?	1	2	3	4	5
25	Quão satisfeito(a) você está com o seu meio de transporte?	1	2	3	4	5

As questões seguintes referem-se a **com que freqüência** você sentiu ou experimentou certas coisas nas últimas duas semanas.

		nenca	alguns veres	freqüentemente	muio freqüentemente	sempre
26	Com que freqüência você tem sentimentos negativos tais como mau humor, desespero, ansiedade, depressão?	1	2	3	4	5

Alguém lhe ajudou a preencher este questionário?.....

Quanto tempo você levou para preencher este questionário?.....

Você tem algum comentário sobre o questionário?

OBRIGADO PELA SUA COLABORAÇÃO