

Daniel Mendes Moreira da Serra

**Avaliação da performance física de atletas
de futebol profissional: um estudo baseado
em dados extraídos de GPS**

Niterói - RJ, Brasil

4 de agosto de 2023

Daniel Mendes Moreira da Serra

**Avaliação da performance física de
atletas de futebol profissional: um
estudo baseado em dados extraídos
de GPS**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Douglas Rodrigues Pinto

Co-Orientador(a): Profa. Dra. Karina Yuriko Yaginuma

Niterói - RJ, Brasil

4 de agosto de 2023

Daniel Mendes Moreira da Serra

**Avaliação da performance física de atletas de
futebol profissional: um estudo baseado em
dados extraídos de GPS**

Monografia de Projeto Final de Graduação sob o título “*Avaliação da performance física de atletas de futebol profissional: um estudo baseado em dados extraídos de GPS*”, defendida por Daniel Mendes Moreira da Serra e aprovada em 4 de agosto de 2023, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Douglas Rodrigues Pinto
Departamento de Estatística – UFF

Profa. Dra. Karina Yuriko Yaginuma
Departamento de Estatística – UFF

Prof. Dr. Marco Aurélio dos Santos Sanfins
Departamento de Estatística – UFF

Prof. Dr. Reinaldo Castro Souza
Departamento de Engenharia Industrial – PUC

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

S487a Serra, Daniel Mendes Moreira da
Avaliação da performance física de atletas de futebol
profissional: um estudo baseado em dados extraídos de GPS /
Daniel Mendes Moreira da Serra. - 2023.
40 f.: il.

Orientador: Douglas Rodrigues Pinto.
Coorientador: Karina Yuriko Yaginuma.
Trabalho de Conclusão de Curso (graduação)-Universidade
Federal Fluminense, Instituto de Matemática e Estatística,
Niterói, 2023.

1. Futebol. 2. Performance física. 3. Pré-processamento de
dados. 4. Análise de Componentes Principais. 5. Produção
intelectual. I. Pinto, Douglas Rodrigues, orientador. II.
Yaginuma, Karina Yuriko, coorientadora. III. Universidade
Federal Fluminense. Instituto de Matemática e Estatística.
IV. Título.

CDD - XXX

Resumo

O presente estudo tem como objetivo analisar a performance física de jogadores profissionais de um clube brasileiro de futebol, a fim de compreender o desempenho nos treinamentos e jogos oficiais. A análise consiste na coleta de dados de dispositivos GPS conectados a cada jogador, tanto durante os treinos como nas partidas oficiais. A partir dessas informações, buscou-se relacionar o esforço físico realizado nos treinamentos com o desempenho físico nos jogos. Para alcançar esse objetivo, foram utilizadas análises de Componentes Principais (PCA) no sentido de compreender as métricas do jogo, e algoritmos de Aprendizado de Máquinas para relacionar os dados dos treinamentos com os jogos. Tanto no PCA quanto no modelo de predição, foram obtidos resultados suficientes, o que permitiu verificar que variáveis relacionadas à velocidade e treinamentos realizados imediatamente um dia antes das partidas desempenham um papel importante no esforço físico dos atletas durante esses jogos.

Palavras-chave: Futebol. Performance Física. Pré-processamento de Dados. Análise de Componentes Principais. Aprendizado de Máquinas. XGBoost.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 9
1.1	Motivação	p. 9
1.2	Objetivos	p. 9
1.3	Revisão Bibliográfica	p. 10
2	Materiais e Métodos	p. 11
2.1	Banco de Dados	p. 11
2.1.1	Pré-processamento	p. 12
2.1.2	Seleção de Atletas	p. 15
2.2	Análise de Componentes Principais (PCA)	p. 18
2.3	Modelos de Aprendizado de Máquinas	p. 19
2.3.1	Regressão Linear	p. 20
2.3.2	<i>Support Vector Machine</i> (SVM)	p. 20
2.3.3	<i>Extreme Gradient Boosting</i> - XGBoost	p. 21
2.3.4	Medidas de Qualidade de Ajuste	p. 22
3	Resultados	p. 24
3.1	Resultados PCA	p. 24
3.2	Análise comparativa entre os modelos	p. 26

3.3 Resultados XGBoost	p. 29
4 Conclusão	p. 32
Referências	p. 34
Apêndice 1 – Descrição das variáveis	p. 36

Lista de Figuras

1	Frequência de treinos no mesmo dia	p. 15
2	Exemplo de Árvore de Decisão	p. 21
3	Força das Correlações Ordenada pela Carga na PC1	p. 25
4	Box-plots das métricas de avaliação para cada modelo	p. 28
5	Gráfico de dispersão entre os valores previstos e reais - com linha de referência	p. 30
6	Gráfico de barras das 10 variáveis mais importantes para o modelo . . .	p. 31

Lista de Tabelas

1	Variáveis Removidas	p. 14
3	Perfil dos jogadores	p. 16
4	Relação da quantidade mínima de treinos antes de um jogo	p. 17
5	Variação Explicada pela PC1 - Método por Posição	p. 26
6	Variação Explicada pela PC1 - Método por Jogador	p. 26
7	Algoritmos testados por categoria	p. 27
9	Tempo de execução dos modelos (segundos)	p. 29
10	Métricas de avaliação por biblioteca	p. 29

1 Introdução

1.1 Motivação

A indústria de esportes não só no Brasil como no mundo, movimenta bilhões de dólares todos os anos. Os principais clubes brasileiros arrecadaram - em 2020 - mais de 4.5 bilhões de reais (Flavio Souza, 2021); conseqüentemente, as equipes têm procurado melhorar a atuação para vencer o maior número de campeonatos possíveis. Junto a isso, muito tem-se estudado sobre o bem-estar dos jogadores e em como otimizar seu desempenho, mas ao mesmo tempo procurando minimizar os riscos físicos que o excesso de treino pode causar, porque além do atleta ficar indisponível por um período de tempo, que pode ser curto ou muito longo, o tratamento físico traz custos altíssimos para o clube. Com o avanço das tecnologias, os tipos de equipamentos que possibilitam extrair dados para analisar e potencializar a performance dos jogadores são inúmeros. Sistemas de Posicionamento Global (GPS) são um dos equipamentos utilizados no monitoramento de treinos Unisport Brasil (2020). Esses dispositivos extraem variáveis como aceleração, velocidade e distância percorrida pelos atletas, permitindo criar diversos tipos de análises que visem o ganho de performance do jogador.

1.2 Objetivos

O objetivo do estudo em debate é realizar uma análise de desempenho físico sobre jogadores profissionais de futebol em partidas oficiais, tendo como base o esforço exercido nos treinos anteriores aos jogos. O primeiro passo é selecionar algumas métricas, obtidas através de dispositivos GPS que são conectados aos atletas, como por exemplo, velocidade máxima ou distância total percorrida, provenientes dos treinos precedentes às partidas e com isso, medir a atuação física do profissional no jogo em evidência. A investigação também dispõe-se a entender quais variáveis influenciam majoritariamente o desempenho físico dos atletas, com a finalidade de auxiliar a criação de treinos que otimizem (ou

potencializem) a atuação física dos profissionais nas partidas, porém, eludindo o aumento do risco de lesões.

1.3 Revisão Bibliográfica

Sucederam uma gama de estudos científicos que buscaram relacionar a performance física do atleta com a quantidade de carga ou percepção de esforço dos treinamentos. Rago et al. (2019) traz uma sumarização e avaliação de diversos estudos desse tipo.

Foi evidenciado em Rago et al. (2020) que a aptidão cardiorrespiratória dos atletas aumenta ao longo de uma temporada e ainda, verificado que existe relação entre esse acréscimo e a distância total percorrida nas sessões de treino. Em contrapartida, diante de Malone et al. (2017), o estudo com o propósito de verificar a associação entre a corrida em alta velocidade (CAV), corridas com arrancada (CA) e lesões concluiu-se que os jogadores que completaram distâncias moderadas em CAV e CA tiveram risco de lesões reduzidos quando comparados aos que completaram baixas distâncias de CAV e CA. Ainda, mostrou que jogadores com mudanças semanais brutas nas distâncias de CAV e CA também tinham o risco de lesão elevado. Isso indica a importância do balanceamento da carga de treino quanto à performance física do jogador. Um ponto importante também encontrado em Malone et al. (2017) é que os atletas que exerceram uma carga de treino crônica tinham risco de lesão reduzidos quando cobriam distâncias de 701 a 750 metros em uma semana de CAV comparados aos que percorreram distâncias menores. Desse modo, fica claro a dificuldade quanto à escolha dos tipos de treinos para que os atletas consigam performar cada vez melhor e ao mesmo tempo não incrementar o risco de lesão. Sendo assim, em Rossi et al. (2018) é apresentado um modelo para prever lesões em jogadores profissionais de futebol. A proposta era utilizar uma abordagem multidimensional para poder prever quando um jogador iria se lesionar, a partir de métricas extraídas dos GPS desses atletas. Após testar diversos algoritmos de *Machine Learning*, o autor chegou a conclusão de que uma Árvore de Decisão seria a melhor abordagem, permitindo assim extrair e investigar desse modelo os principais motivos que levam às lesões observadas. O que consequentemente seria um auxílio para os treinadores e técnicos, permitindo assim, melhores tomadas de decisão quanto à realização de sessões de treino.

2 Materiais e Métodos

2.1 Banco de Dados

Para o trabalho foram utilizados dados de jogadores profissionais de um clube brasileiro da 1^a divisão do Campeonato Brasileiro no ano de 2019. A base de dados utilizada foi criada a partir de 326 arquivos do tipo csv, onde cada um desses arquivos equivalia a um jogo ou uma sessão de treino por dia. Em alguns casos foi possível encontrar mais de um arquivo com a mesma data, esses casos se tratavam de jogadores que, por algum motivo, não estavam presentes no arquivo principal do respectivo dia. Então, gerou-se um novo csv para esses atletas. Todos os arquivos apresentavam a mesma estrutura, ou seja, as mesmas colunas com os mesmos nomes, organizadas na mesma ordem. Essas colunas exibiam informações obtidas através de equipamentos GPS, que eram conectados um em cada jogador em cada sessão de treino ou em uma partida oficial. Esses dispositivos forneciam métricas como aceleração, velocidade média, velocidade máxima, entre outras. Além dessas variáveis, também estavam presentes na base de dados informações referentes às características do atleta como data de nascimento, peso e altura. Só apareciam no arquivo de treino os atletas que, de fato, participaram daquela sessão, assim como nos arquivos de jogos. Foram encontrados 52 arquivos de sessões de treino contendo colunas com nomes distintos das dos demais. Essas colunas continham informações que no padrão dos demais arquivos, estavam localizados em colunas com outros nomes. As colunas divergentes eram as seguintes:

- Coluna “Distance Zone 4” continha informações da coluna padrão “Distance Zone 4 (Relative)”
- Coluna “Distance Zone 5” continha informações da coluna padrão “Distance Zone 5 (Relative)”
- Coluna “Distance Zone 6” continha informações da coluna padrão “Distance Zone 6 (Relative)”

- Coluna “High Speed Running” continha informações da coluna padrão “High Speed Running (Relative)”
- Coluna “HSR Per Minute” continha informações da coluna padrão “HSR Per Minute (Relative)”
- Coluna “Max Acceleration” continha informações da coluna padrão “Max Acceleration (Disabled)”
- Coluna “Max Deceleration” continha informações da coluna padrão “Max Deceleration (Disabled)”

O tratamento realizou-se substituindo os valores, que eram faltantes, das colunas padrão pelos valores contidos nas colunas divergentes apenas nos casos desses 52 arquivos de treino. Após, as colunas divergentes foram excluídas, sendo possível unificá-las em uma tabela. A base final era composta por observações que representavam os jogadores do sexo masculino em um dado treinamento ou uma partida oficial, com as variáveis associadas aos mesmos. No total haviam 216 dias de treino, 57 dias de jogos de 3 campeonatos nacionais distintos e 59 jogadores com um total de 215 variáveis. A descrição de todas as variáveis pode ser encontrada no Apêndice 1.

2.1.1 Pré-processamento

O pré-processamento dos dados desempenha um papel fundamental na preparação de qualquer conjunto de dados para análise. Neste trabalho, foram realizadas várias etapas de limpeza e organização a fim de garantir a qualidade e a consistência dos dados utilizados.

Com a importação da base para o ambiente do RStudio, uma das primeiras etapas foi a padronização do formato dos campos de data. Os dados originais continham diferentes representações de datas, o que poderia afetar a análise posterior. Portanto, as variáveis de data “Drill Date” e “Session Date” foram padronizadas para o formato dd/mm/YYYY. Além disso, as colunas “Player Position” e “Drill Title” que traziam, respectivamente, a posição do jogador e o título do exercício realizado, também passaram por um processo de padronização. Como eram campos do tipo *string*, um dos passos para evitar duplicidade de informação é não diferenciar caracteres minúsculos de maiúsculos, com isso, ambas as variáveis foram padronizadas para o formato minúsculo. E ainda, “Player Position” que é uma variável categórica, foi convertida para o tipo fator para facilitar a análise. Da

mesma forma, a variável “categ” que foi criada a fim de categorizar dada sessão, ou seja, caso o exercício respectivo seja referente a um jogo de um campeonato específico, essa coluna retorna o nome do campeonato; caso o exercício seja um treino, o retorno será “treino”, também foi convertida para fator.

Durante o pré-processamento dos dados a criação variável “Drill Duration (min)” fez-se necessária. A partir das informações das variáveis “Drill End Time” e “Drill Start Time” essa nova coluna foi inserida ao banco de dados representando o tempo total da sessão de treino/jogo em minutos. E logo notou-se *outliers* nesta variável. Treinos e jogos com mais de 180 minutos de duração ou, até mesmo, uma duração negativa. Tais valores foram considerados possíveis erros, acionamentos precoces ou desligamentos tardios dos aparelhos de GPS. Portanto, optou-se por excluir as observações que continham esses valores discrepantes, a fim de garantir a integridade dos dados e evitar distorções nos resultados.

Outra etapa importante foi a padronização dos nomes dos jogadores. Observou-se que os nomes dos atletas eram escritos de formas diferentes em algumas instâncias, as vezes com algum erro de codificação no texto, escritos de forma incorreta ou contendo sobrenomes diferentes. Com o objetivo de garantir a consistência e evitar ambiguidades, os nomes foram padronizados de forma única, isso facilitou a identificação correta dos jogadores e a análise subsequente.

Ainda no processo de tratamento, fez-se necessária a exclusão de algumas variáveis que não iriam agregar a análise. Na Tabela 1 encontram-se as variáveis removidas e a justificativa para remoção em cada grupo.

Durante a exploração dos dados, constatou-se a presença de linhas duplicadas. A fim de evitar vieses e inconsistências nos resultados - essas linhas foram removidas - mantendo apenas uma única ocorrência de cada registro.

Uma coluna relevante nos dados era o “Drill Title”, que representava o título da respectiva sessão de treino ou jogo. No entanto, esse campo não possuía um padrão definido. Eram 427 valores únicos presentes. Para facilitar a análise e agrupamento dessas informações, realizou-se a renomeação do máximo de observações possíveis para categorias que representavam o tipo de exercício efetuado nos treinos. O número final de valores distintos passou a ser 81, com as dez primeiras categorias acumulando 76% da frequência. Essa padronização permitiu uma análise mais consistente e uma melhor compreensão dos resultados.

Tabela 1: Variáveis Removidas

Grupo de Variáveis	Motivo da Exclusão
“% Time in Red Zone”, “Custom Player Id”, “Custom Metric X”, “Decelerations Total Distance Zx”, “Entries Zone x (Absolute) (Disabled)”, “Entries Zone x (Relative) (Disabled)”, “Entries Zone x (Absolute)”, “Entries Zone x (Relative)”, “Heart Rate Exertion”, “Heart Rate Recovery Beats”, “Heart Rate Recovery Percent”, “Heart Rate Variability”, “Nutrition”, “Osmolarity”, “Perceived Exertion”, “Readiness”, “Sleep”, “Stress”, “Subjective Q6”, “Time In Heart Rate Zx to Zy”, “Time In Heart Rate ZoneX” e “Time In Red Zone”	Zeradas, maioria das observações com 0 ou nulas
“Accelerations Total Distance Z1”, “Accelerations Total Distance Z2”, “Accelerations Total Distance Z3”, “Acute Vs Chronic Ratio”, “Average Heart Rate”, “Body”, “Duration of High Intensity Bursts”, “Max Acceleration”, “Max Deceleration”, “Max Heart Rate” e “Speed Intensity Zone6”	Variância zero ou próxima de zero
“Distance ZxToZy”, “Distance Zone X (Absolute)”, “High Speed Running (Relative)” e “HSR Per Minute (Relative)”	Foi mantido o absoluto ou relativo dependendo da variável excluída
“Distance Zonal”, “Drill Date”, “Drill End Time”, “Drill Id”, “Drill Start Time”, “Player Display Name”, “Player First Name”, “Player Last Name”, “Session End Time” e “Session Start Time”	Substituídas por variáveis equivalentes
“HML Time”, “Interval”, “Metabolic Time Zonal”, “Metabolic Time Zone x”, “Session Day of Week”, “Session Title”, “Session Type”, “Session Week Number”, “Time Zone x (Absolute)”, “Time Zone x (Relative)”, “Total Time” e “Player Date Of Birth”	Desnecessárias ou inconsistentes

Identificou-se mais de uma observação de treino para o mesmo jogador no mesmo dia. A Figura 1 mostra o gráfico de frequência da quantidade de treinos no mesmo dia. Logo, entendeu-se a necessidade de, posteriormente, unir essas métricas, agrupando os dados por jogador e dia de treino e assim, podendo somar, realizar médias simples ou ponderadas ou obter os valores máximos para cada métrica no agrupamento, dependendo da operação que faria mais sentido para cada uma delas. Além disso, foi identificado que haviam 27 dias de treinos realizados no mesmo dia de jogos. Em linha com o objetivo do estudo, escolheu-se avaliar apenas os treinos realizados em dias anteriores aos jogos, de modo a desconsiderar os que não se encaixavam nesse critério.

No caso dos jogos, observou-se que o campo “Drill Title” geralmente indicava o primeiro e segundo tempo do jogo ou até mesmo a sessão inteira. No entanto, também foram encontrados outros tipos de títulos que não eram relevantes para a análise como aquecimento, por exemplo. Portanto, optou-se por excluir esses registros específicos, focando apenas nos dados pertinentes aos jogos.

Um critério adotado para a exclusão de observações inconsistentes foi a verificação do total de colunas iguais a zero. Observou-se que algumas linhas continham uma quantidade significativa de colunas com valor zero, o que poderia indicar problemas na coleta dos dados ou informações ausentes. Assim, excluiu-se as linhas em que a quantidade

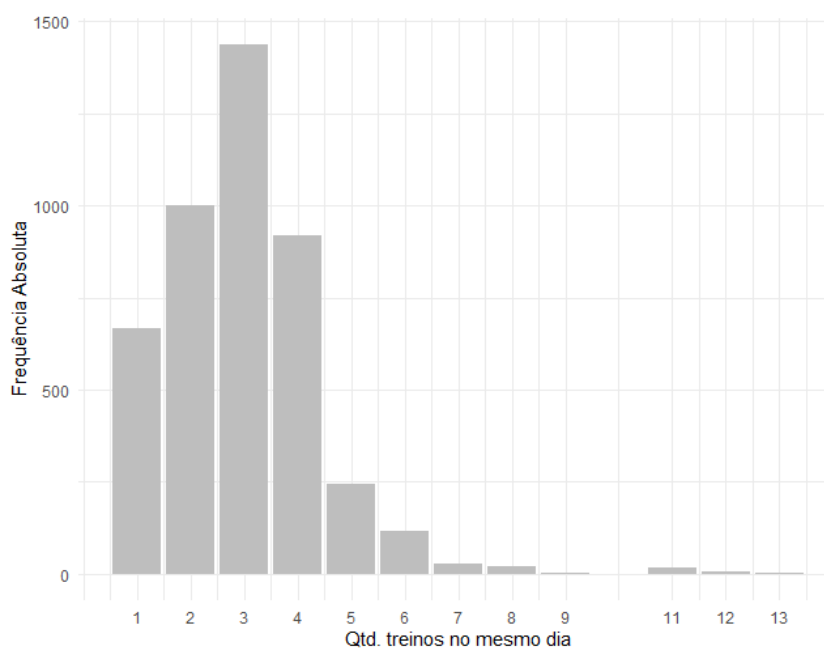


Figura 1: Frequência de treinos no mesmo dia

de variáveis iguais a zero era superior a metade do total de colunas, visando manter a consistência e a confiabilidade dos dados analisados.

Essas etapas de limpeza foram executadas com o intuito de garantir a qualidade e a consistência dos dados utilizados neste trabalho. Ao realizá-la, buscou-se obter resultados mais precisos e confiáveis na análise.

2.1.2 Seleção de Atletas

A seleção dos atletas aptos para o estudo é uma etapa crucial. Nesta subseção, descreve-se as estratégias adotadas para selecionar os atletas com base nas informações de escalação.

Para obter informações precisas sobre quais jogadores participaram efetivamente dos jogos, foi utilizada a API SofaSport, do site Sofascore que agrega dados como escalações e estatísticas sobre diversos campeonatos e diversos times, tanto de futebol quanto de outros esportes. Através da linguagem Python foi possível criar um notebook na plataforma Google Colab e extrair as escalações referentes aos jogos disponíveis. Essa abordagem permitiu selecionar apenas os jogadores que estavam presentes e atuaram nos jogos, evitando qualquer viés de inclusão de quem não participou das partidas analisadas. Conseguir as escalações através da API foi suficiente para dois dos três campeonatos presentes no banco de dados, mas para o faltante foi necessário encontrar outra fonte, visto que deta-

lhes desse campeonato não estavam contemplados no Sofascore. As escalações faltantes foram extraídas manualmente do site da Federação Carioca de Futebol, concluindo assim todas as informações de participações possíveis.

Com a inclusão das escalações no banco de dados analisado, conseqüentemente incluiu-se duas novas variáveis.

- `jogou`: Variável indicadora que mostra se o atleta jogou dada partida (1) ou não (0);
- `round_info`: Variável que informa a rodada do campeonato respectivo.

Algumas características dos profissionais analisados estão consolidadas na Tabela 3.

Tabela 3: Perfil dos jogadores

Características	Média	Desvio-padrão
Idade	25.93	4.23
Altura	180.13	5.50
Peso	78.47	7.21

A fim de estudar a relação dos treinos com os jogos, foi necessário verificar quantos treinos um jogador participava antes de jogar efetivamente uma partida. Para isso, foi utilizada uma função que contabilizava a quantidade de treinos em que um atleta havia atuado antes de um jogo em que ele tenha de fato participado. No entanto, devido à limitação das informações disponíveis, não haviam registros de todos os jogos da temporada. Com isso, ao desenvolver a função para contar os treinos antes de um jogo, também considerou-se as datas ausentes, uma vez que desconsiderar essas datas impactaria em uma contagem errônea.

Como resultado desse processo, uma nova tabela foi consolidada para averiguar a quantidade de vezes em que um jogador obteve no mínimo 2, 3, 4 e 5 treinos antes de uma partida em que ele tenha sido escalado. A Tabela 4, com as seguintes colunas, mostra essa relação para cada atleta:

- `Player ID` = Identificação única do jogador;
- `min_2` = Quantidade de vezes em que o atleta participou de, pelo menos, 2 treinos antes de jogar uma partida oficial;
- `min_3` = Quantidade de vezes em que o atleta participou de, pelo menos, 3 treinos antes de jogar uma partida oficial;

- min_4 = Quantidade de vezes em que o atleta participou de, pelo menos, 4 treinos antes de jogar uma partida oficial;
- min_5 = Quantidade de vezes em que o atleta participou de, pelo menos, 5 treinos antes de jogar uma partida oficial;

Tabela 4: Relação da quantidade mínima de treinos antes de um jogo

Player ID	min_2	min_3	min_4	min_5
X01	9	8	6	4
X02	19	15	7	3
X03	5	2	2	1
X04	3	1	1	1
X05	27	17	13	7
X06	9	6	6	4
X07	12	9	7	6
X08	3	2	1	1
X09	18	13	9	5
X10	20	12	8	4
X11	4	4	2	0
X12	5	1	0	0
X13	4	2	1	1
X14	5	4	3	2
X15	15	11	9	6
X16	30	20	14	7
X17	14	11	6	5
X18	16	11	6	3
X19	8	4	3	2
X20	14	9	7	5
X21	9	7	6	3
X22	21	15	10	4
X23	7	6	5	5
X24	6	4	3	1
X25	11	9	6	4
X26	14	8	6	2
X27	25	13	11	7
X28	10	4	3	2

Com base nesses dados, o critério para a avaliação de jogos foi de haver ao menos 3 treinos realizados pelo profissional antes de tal partida. E ainda, o jogador deveria ter ao menos 8 ocorrências de no mínimo 3 treinos antes de um jogo (min_3), para minimizar o risco de inconsistências em cada profissional.

Para legitimar os critérios de seleção notou-se a distribuição do número de participantes dos artigos presentes em Rago et al. (2019), onde o autor examina diversos estudos em

que recorreu-se diferentes tipos de tecnologia para coletar e interpretar a carga de treino de jogadores profissionais, e constatado se o número de atletas aptos estava em linha com essa distribuição. Como em Rago et al. (2019) o número mínimo de participantes era 6 e o 1º quartil apresentava 17 participantes, a quantidade escolhida de 15 candidatos pertencia a essa distribuição.

2.2 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais (PCA) é um método utilizado para reduzir variáveis do conjunto de dados analisado, visto que estima uma componente principal para cada uma das variáveis pertencentes ao banco de dados, onde cada componente é uma combinação linear de todas essas, ou seja, individualmente as componentes carregam informações sobre todas as colunas.

O PCA considera as variáveis X_1, X_2, \dots, X_n , presentes no conjunto de dados com p observações. E a partir disso, a matriz X de ordem $p \times n$ representa esses dados.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

Dado a matriz X é possível estimar uma nova matriz de variâncias e covariâncias S de ordem $n \times n$.

$$S = \begin{bmatrix} Var(x_1) & Cov(x_1x_2) & \cdots & Cov(x_1x_n) \\ Cov(x_2x_1) & Var(x_2) & \cdots & Cov(x_2x_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_nx_1) & Cov(x_nx_2) & \cdots & Var(x_n) \end{bmatrix}$$

No PCA é importante, na maior parte das vezes, padronizar a escala das suas variáveis, dado que a não padronização pode afetar significativamente a análise. Sendo assim, a padronização é feita considerando média zero e variância um, onde \bar{X}_j e S_{x_j} representam, respectivamente, a estimativa da média e desvio padrão da coluna j .

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_{x_j}}, i = 1, 2, \dots, p \text{ e } j = 1, 2, \dots, n \quad (2.1)$$

Após a padronização, a matriz Z é obtida.

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{p1} & z_{p2} & \cdots & z_{pn} \end{bmatrix}$$

A partir da matriz S , é possível obter a equação característica que tem como autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$. Onde, para todo λ_n , λ_{n-1} é maior que λ_n . E para cada autovalor existe um autovetor \tilde{a}_i , onde:

$$\tilde{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix}$$

Com isso, pode-se obter as componentes principais. Sendo a i -ésima dada por:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{in}X_n \quad (2.2)$$

Considerando as componentes principais, é possível calcular a carga C_i de cada componente, ou seja, quantos por cento cada uma contribuí para explicar a variância total.

$$C_i = \frac{\hat{Var}(Y_i)}{\sum_{i=1}^n \hat{Var}(Y_i)} \cdot 100 \quad (2.3)$$

Dessa forma pode-se avaliar quantos por cento da variação é explicado por k componentes Carlos Alberto Alves Varella (2008).

2.3 Modelos de Aprendizado de Máquinas

A seguinte seção descreve alguns métodos de aprendizado de máquinas utilizados na análise, evidenciando sua lógica, vantagens e desvantagens. Ainda, reúne informações sobre medidas de qualidade de ajuste dos modelos que serão utilizadas como avaliações do resultado da implementação de cada um dos algoritmos, possibilitando comparar a performance desses.

2.3.1 Regressão Linear

Existem diversas técnicas utilizadas para prever o valor de uma variável no aprendizado de máquinas. No presente trabalho testou-se algumas dessas metodologias, dentre elas a regressão linear. Uma regressão linear consiste em associar uma ou mais variáveis independentes a uma variável dependente onde, para cada uma das variáveis independentes estima-se um coeficiente compondo assim a chamada equação linear. A partir dessa equação é possível obter-se um valor para a métrica resposta.

O cálculo realizado na regressão linear está representado na equação 2.4, onde Y representa a variável resposta que deseja-se prever ou explicar, X_i a i -ésima variável independente, β_0 o coeficiente linear, β_i o coeficiente angular para a i -ésima variável, que representa o efeito que cada variável independente exerce na dependente e e equivale ao termo de erro da equação que reflete toda a variação não explicada pelo modelo.

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \dots + \beta_i.X_i + e \quad (2.4)$$

Ao implementar um modelo de regressão linear é preciso atentar-se aos pressupostos desse método. As variáveis precisam ter uma relação linear, ou seja, a variável Y deve variar de acordo com o acréscimo ou decréscimo da variável X de uma forma constante, dado que relações não-lineares ou muito fracas indicam que o modelo não irá funcionar de forma satisfatória. Além disso, os resíduos devem apresentar um comportamento independente e homocedástico dado que, a existência de algum padrão nesses erros indica que o modelo não está capturando todos os fatores que influenciam na variação da variável dependente. Além disso o conjunto de dados utilizado na predição deve possuir distribuição Normal, visto que modelos de regressão são muito afetados por dados assimétricos o que acaba prejudicando a assertividade do algoritmo João Guilherme Berti Sczip (2019).

2.3.2 *Support Vector Machine* (SVM)

Uma outra abordagem para realizar a predição de dados contínuos são os modelos SVM (*Support Vector Machine*). É um método robusto que pode ser utilizado tanto em problemas de regressão como de classificação Bernardo Coutinho (2019). Esse algoritmo consiste em encontrar um hiperplano que separe os dados de forma que a margem seja a maior possível. Assim, os dados mais próximos do hiperplano, vetores de suporte, são utilizados para treinar o modelo. A equação 2.5 representa uma regressão SVM, sendo a_i

e b os parâmetros que minimizam a função de perda 2.6 e $K(x_i, x)$ o *kernel* que constitui em medir a semelhança entre os pontos x_i e x . Existem diversos tipos de *kernel* que podem ser utilizados na aplicação de um modelo SVM, como linear, polinomial e radial.

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad (2.5)$$

$$L(\alpha_1, \dots, \alpha_n, b) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{i=1}^n \alpha_i^2 \quad (2.6)$$

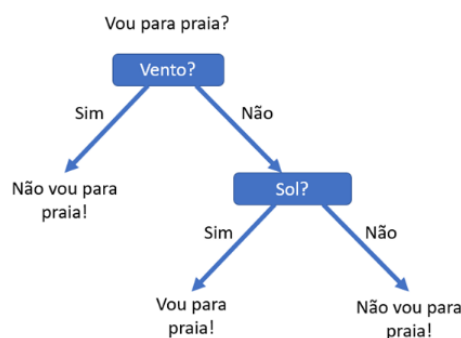
A função de perda apresenta o hiperparâmetro λ que controla a complexidade do modelo. Quanto maior o λ mais complexo é o modelo, ou seja, maior a capacidade de ajuste aos dados, porém torna-se mais propenso a adaptar-se aos ruídos presentes. Quanto menor o λ mais simples é o modelo, ou seja, menor a capacidade de ajuste, porém a propensão de adaptar-se a um viés também diminui.

Algoritmos SVM são eficientes computacionalmente e aptos a lidar com dados não-lineares a partir do *kernel*. Contudo, são mais complexos, tanto para entendimento quanto para a implementação correta. Precisam do ajuste correto de vários hiperparâmetros, o que pode ser uma grande desvantagem, dadas as complicações de compreensão desses.

2.3.3 *Extreme Gradient Boosting - XGBoost*

XGBoost é um modelo de aprendizado de máquinas baseado em árvores de decisão, que são mapas de diversas escolhas possíveis a serem selecionadas. A Figura 2 demonstra bem um exemplo de estrutura de uma árvore de decisão.

Figura 2: Exemplo de Árvore de Decisão



Fonte: (Didática Tech, 2022)

O XGBoost combina o resultado de diversas árvores de decisão para chegar na melhor

predição possível, pois para cada modelo de árvore criado, o modelo seguinte tem como objetivo corrigir os equívocos do modelo anterior.

O algoritmo apresenta a função de perda 2.7, que é chamada de função de perda do erro quadrático. Onde y_i é o valor observado e $f(x_i)$ é o valor previsto.

$$L(f(x), y) = \frac{1}{2} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2.7)$$

O XGBoost busca minimizar a função de perda ajustando os parâmetros de forma iterativa. Ele começa o processo com conjuntos aleatórios de parâmetros e busca sempre minimizar a função L até que essa não possa mais ser otimizada.

Essa abordagem apresenta diversas vantagens que até melhoram seu desempenho, como a paralelização que ajuda a reduzir o tempo de processamento de treino do modelo Onédio Junior (2022). Também permite lidar com valores ausentes, o que reduz a necessidade de tratar esses valores. Além disso, tem um ajuste automatizado caso os parâmetros não sejam ajustados da forma correta, acarretando em um resultado satisfatório apesar dessa configuração errônea.

Mas também apresentam pontos negativos, como o maior tempo de processamento em comparação com outros modelos, visto que o processo de minimizar o erro acaba sendo muito exaustivo. Sensível a valores extremos, não funciona corretamente com dados não estruturados e basear a correção em estimadores anteriores acaba implicando em um método não escalável.

2.3.4 Medidas de Qualidade de Ajuste

Para entendermos o desempenho do modelo, é necessário checar algumas métricas de avaliação. As duas levadas em consideração na análise foram o Coeficiente de Determinação (R^2) e a Raiz do Erro Quadrático Médio (RMSE).

O R^2 é calculado a partir dos valores reais da variável dependente (y), os valores preditos para esta variável (\hat{y}) e a média dos valores reais (\bar{y}).

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.8)$$

Através da Equação 2.8 é possível obter o Coeficiente de Determinação, que indica o percentual da variância da variável dependente que é explicado pelo modelo. Quanto

mais alto for o R^2 , mais a métrica predita é consequência das independentes.

A Raiz do Erro Quadrático Médio é uma métrica variante do Erro Quadrático Médio (MSE) que faz o cálculo da média de diferença entre o valor predito e o real, elevando essa diferença ao quadrado, o que penaliza valores com uma diferença extrema Clébio de Oliveira Júnior (2021).

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.9)$$

Porém, elevar esse erro ao quadrado acaba trazendo um viés na interpretabilidade da métrica, pois acaba alterando a unidade de medida da variável predita. Com isso, aplica-se a raiz quadrada para tornar a escala da variável original, sendo esta aplicação da raiz a criação da RMSE.

$$RMSE(y, \hat{y}) = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.10)$$

Existe, ainda, o Erro Absoluto Médio (MAE) que mede a diferença média entre os valores previstos e os valores reais, calculado na equação 2.11.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.11)$$

O MAE não faz suposições sobre a distribuição dos dados, ou seja, é uma medida não-paramétrica e não é afetada por valores extremos.

Essas métricas são essenciais para entender a robustez do modelo. Caso não apresentem resultados satisfatórios, pode-se tomar a decisão de alterar a abordagem ou realizar pequenos ajustes que resultem em uma performance mais aceitável. Na circunstância de indicar que o modelo funciona de forma eficiente, servem como validação da predição configurada podendo ainda, mesmo com uma performance suficiente, revelar algum ponto de aperfeiçoamento do algoritmo.

3 Resultados

3.1 Resultados PCA

Implementando o PCA no estudo em questão, o objetivo principal foi obter um *score* que representasse de forma consistente a performance física do atleta na partida disputada. E para montar esse *score*, o PCA foi o método que melhor se adequou, pois apresenta o propósito de resumir as variáveis em componentes principais. O objetivo foi utilizar a componente principal 1 (PC1) como *score*, caso a carga (peso) dessa fosse suficiente para prosseguir com as análises.

O primeiro passo para testar o algoritmo de componentes principais no banco de dados, foi considerar apenas variáveis numéricas. Portanto, as colunas “Player Name”, “Session Date”, “Drill Title”, “Player Position”, “categ”, “round_info” foram removidas, assim como outras 5 que apresentavam 3 motivos diferentes para serem desconsideradas:

- “jogou”: Coluna indicadora que informa se o atleta participou ou não do jogo. Como a análise considera apenas atletas escalados, então essa coluna é constante para todos;
- “Player Id” e “Session Id” = Variáveis de identificação única, de jogador e sessão de exercício, respectivamente. Apesar de estarem configuradas como numéricas, não trazem nenhuma informação referente a performance física dos jogadores;
- “Player Height” e “Player Weight” = Altura e peso dos atletas, respectivamente. Mesmo sendo informações referentes ao físico dos candidatos, o fato de serem constantes, individualmente, pode acabar retornando algum viés, visto que a análise de componentes principais avalia a variância.

Com isso a função `prcomp` do pacote `stats` que realiza o PCA nos dados, pôde ser devidamente aplicada apenas nas informações de jogo, visto que não precisamos de um

score para os treinos. Evidenciando a configuração do parâmetro *scale*, sendo igual a *True*, indicando que os dados serão normalizados no processo.

O algoritmo resultou em uma proporção de variância na componente principal 1 (PC1) igual a 0.4838, ou seja, a 1^a componente explica 48% das variações presentes nas variáveis originais. Apesar de não ser um valor expressivamente alto, para fins analíticos, ter um *score* que resume, aproximadamente, 50% do que foi exercido em jogo, é uma métrica que traz muito valor e será suficiente para a análise.

Logo, acrescentou-se os valores obtidos para a PC1 de cada observação ao banco de dados. Para checar se os valores da nova coluna estavam em linha com o crescimento e decréscimo das demais realizou-se duas análises de correlação. Uma considerando a metodologia de Pearson e a outra de Spearman, representadas na Figura 3 que mostra as correlações de ambos os métodos para cada variável indexadas em ordem crescente a partir do peso de cada uma dessas métricas na componente principal 1. Como, no gráfico, temos ambas as linhas decrescentes da esquerda para a direita, isso reflete que quanto menor o peso da variável na PC1, mais fraca é a correlação. Um resultado que vai em linha com o esperado para o *score* estimado, ou seja, quanto maior os valores das variáveis que impactam positivamente na performance, maior o *score*.

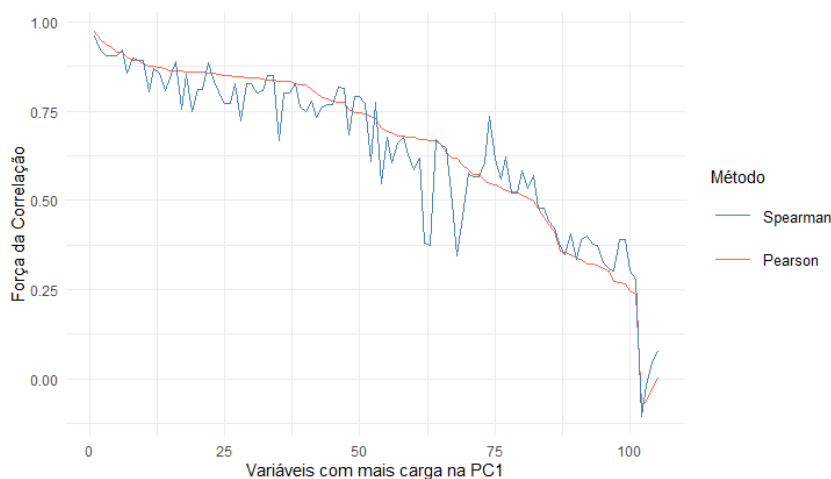


Figura 3: Força das Correlações Ordenada pela Carga na PC1

Também foram realizados testes com outras configurações do PCA. Uma segmentando por jogador, ou seja, um PCA para cada atleta e outra separada por posição. Porém, essas não obtiveram um desempenho satisfatório, o método sem separação foi, de fato, superior. Alguns resultados com o valor da variação explicada pela PC1 no modelo por posição estão consolidados na Tabela 5 e por jogador na Tabela 6.

Tabela 5: Variação Explicada pela PC1 - Método por Posição

Posição	Variação Explicada
Volante	0,52
Meia	0,69
Lateral	0,49
Zagueiro	0,26
Atacante	0,52

Tabela 6: Variação Explicada pela PC1 - Método por Jogador

Player ID	Variação Explicada
X01	0,61
X02	0,69
X05	0,50
X07	0,79
X09	0,34
X10	0,45
X15	0,56
X16	0,61
X17	0,45
X18	0,45
X20	0,35
X22	0,63
X25	0,56
X26	0,39
X27	0,45

3.2 Análise comparativa entre os modelos

A aplicação de um modelo de aprendizado de máquinas na análise em questão, consiste em considerar as variáveis dos 3 treinos anteriores como variáveis explicativas, ou seja, essas serão utilizadas como métricas que explicam a variável resposta. E a variável resposta, que será a prevista por essas métricas, será o *score* calculado pelo PCA no jogo que precede os 3 últimos treinos. Podendo assim, obter-se informações a cerca do desempenho físico do atleta nas partidas a partir das estatísticas coletadas deles através de GPS nos treinos anteriores.

Para realizar a implementação do modelo, antes foi necessário reestruturar o *data-frame*. Até este ponto, os dados estavam a nível de atleta por dia de jogo ou treino sendo as colunas, referentes as métricas físicas. Montou-se uma função para criar a tabela que será utilizada para implementar o algoritmo de predição. Após a aplicação dessa função, a estrutura da nova tabela consistia em nível de atleta por dia de jogo e as métricas para tal partida foram retiradas dos 3 dias de treinos anteriores ao jogo. Ou seja, as colunas eram

compostas por todas as métricas dos 3 treinos anteriores, com um sufixo indicando a qual treino aquela variável pertencia. Por exemplo, se antes tinha-se uma métrica chamada “Acceleration” para os dias t_{-1} , t_{-2} e t_{-3} , sendo t_{-i} o treino referente ao i -ésimo dia antes do jogo, na tabela estruturada para o modelo as 3 “Acceleration” para cada treino estavam consolidadas em uma mesma linha com os sufixos indicando o treino de referência, então, “Acceleration $_{t_{-1}}$ ”, “Acceleration $_{t_{-2}}$ ” e “Acceleration $_{t_{-3}}$ ”. Podendo assim relacionar as métricas dos treinos anteriores com o *score* do jogo em questão, que também fez parte da base estruturada.

A partir de uma comparação com diversos outros modelos, dentre eles, algoritmos de regressão linear e SVM, foi feita a escolha do método XGBoost. No aprendizado de máquinas, comparar o desempenho de diferentes modelos é uma etapa muito importante do processo. Através de uma metodologia chamada *Cross Validation*, foi possível testar e consolidar as métricas de desempenho para cada algoritmo e, com isso, tomar a decisão de utilizar o que melhor respondia a análise. Onze tipos de algoritmos diferentes foram testados. A Tabela 7 mostra as 3 categorias distintas e cada um dos modelos pertencentes a dado grupo.

Tabela 7: Algoritmos testados por categoria

Categoria	Algoritmos
<i>Gradient Boosting</i>	xgbTree, xgbLinear, gbm
Regressão Linear	lm
<i>Support Vector Machine</i>	svmLinear, svmLinear2, svmLinear3, svmPoly, svmRadial, svmRadialCost, svmRadialSigma

O método de validação cruzada chamado *K-fold* foi utilizado nos modelos listados para fins de comparação desses algoritmos. O *K-fold* consiste em dividir o conjunto de dados em treino e teste por x vezes e avaliar o desempenho desses nas n vezes em que processo se repete. Com isso, esse método resulta em uma medida de desempenho menos sujeita ao acaso. Para a seguinte análise foi considerado 10 subdivisões de treino e teste e 3 repetições, ou seja, a validação cruzada dividiu os dados em 10 e cada um dos subconjuntos foi utilizado como conjunto de teste uma vez e os 9 restantes como conjunto de treino, sendo esse processo repetido por 3 vezes usando subconjuntos diferentes.

Levou-se em consideração algumas métricas de desempenho mas principalmente o Coeficiente de Determinação (R^2) e o tempo necessário para a execução completa de

cada um. A Figura 4 mostra a distribuição dos valores obtidos para R^2 de cada um dos modelos.

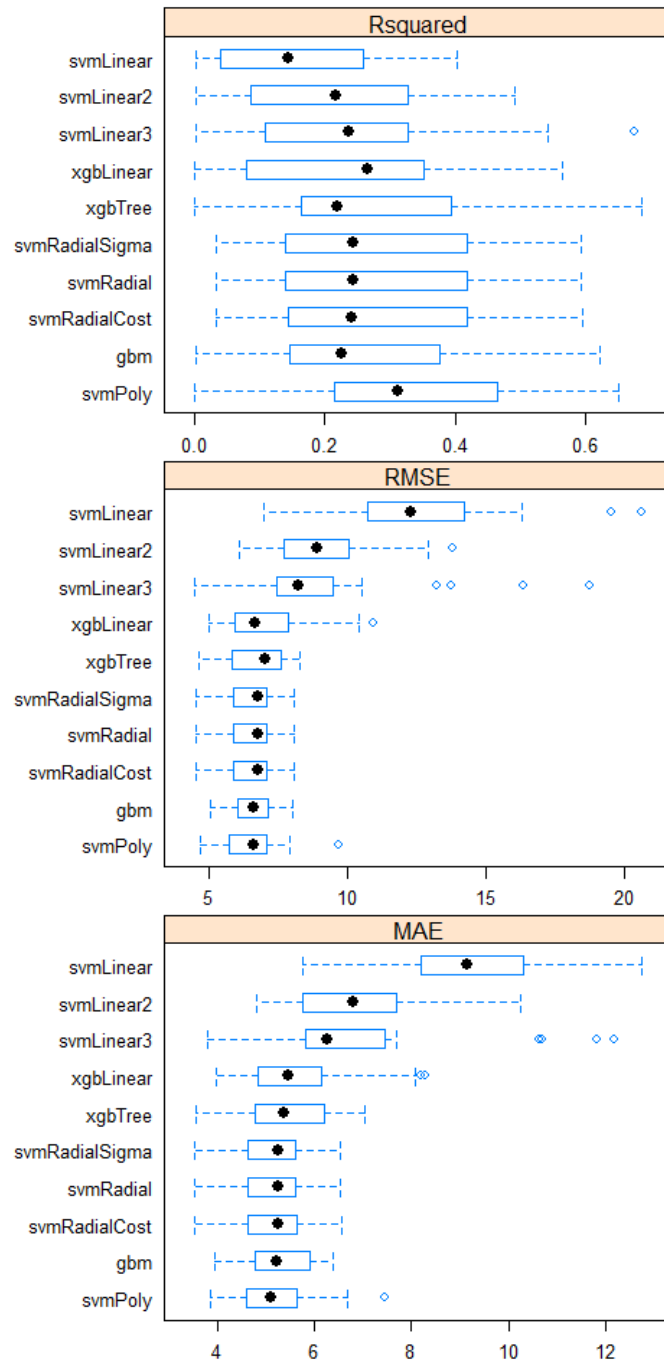


Figura 4: Box-plots das métricas de avaliação para cada modelo

Verificando os valores de R^2 , o `xgbTree` têm uma ampla variação dos valores, porém é bem concentrado entre o 1º e 3º quartis. O `svmPoly` também mostra um desempenho similar, e é o que apresenta melhor mediana. Portanto escolheu-se ambos para serem testados individualmente e ainda, o `gbm`, que obteve um desempenho parecido com o

`xgbTree`. Porém, checando o tempo de execução na Tabela 9, nota-se a velocidade superior ao `xgbTree`.

Tabela 9: Tempo de execução dos modelos (segundos)

Modelo	Tempo Total (seg)
<code>xgbTree</code>	236,48
<code>xgbLinear</code>	229,00
<code>gbm</code>	21,34
<code>svmLinear</code>	5,70
<code>svmLinear2</code>	11,85
<code>svmLinear3</code>	44,58
<code>svmPoly</code>	30,25
<code>svmRadial</code>	3,84
<code>svmRadialCost</code>	3,69
<code>svmRadialSigma</code>	9,91
<code>lm</code>	1,80

Em relação ao RMSE, todos são bem parecidos, com exceção dos 3 `svmLinear` e o `xgbLinear` que têm distribuições que apresentam maiores erros.

Com os 3 candidatos, realizou-se a implementação na amostra de validação (teste). Para isso, cada modelo foi executado através de dois métodos. O 1º por meio da biblioteca computacional `caret`, utilizada para realizar o *Cross Validation* e o 2º por meio dos pacotes específicos para cada um dos três algoritmos. A partir da Tabela 10 é possível notar que tanto o RMSE quanto o R^2 da função `xgb.train` são muito superiores em relação a todas as demais, o que implicou na escolha da biblioteca `xgboost` para prosseguir com a análise.

Tabela 10: Métricas de avaliação por biblioteca

Biblioteca	Função	RMSE	R^2
<code>caret</code>	<code>train(method="xgbTree")</code>	6.769	0.202
<code>xgboost</code>	<code>xgb.train</code>	5.332	0.487
<code>caret</code>	<code>train(method="svmPoly")</code>	7.337	0.203
<code>e1071</code>	<code>svm(kernel="polynomial")</code>	7.206	0.145
<code>caret</code>	<code>train(method="gbm")</code>	6.255	0.295
<code>gbm</code>	<code>gbm</code>	6.459	0.237

3.3 Resultados XGBoost

Após a implementação do algoritmo de aprendizado de máquinas XGBoost, obteve-se um Coeficiente de Determinação (R^2) de 48,7%, ou seja, as variáveis independentes

referentes aos 3 treinos anteriores ao jogo em questão explicam quase 50% do *score* obtido pelo PCA da componente principal 1 que, por sua vez, explica 48,38% das variações encontradas nas métricas relativas aos jogos. A Figura 5 mostra o gráfico de dispersão entre os valores previstos e os valores reais. Nele, é possível perceber que as previsões, em geral, não se distanciam tanto da linha de referência, ou seja, as previsões estão próximas do valor real, indicando que o modelo obteve um bom ajuste.

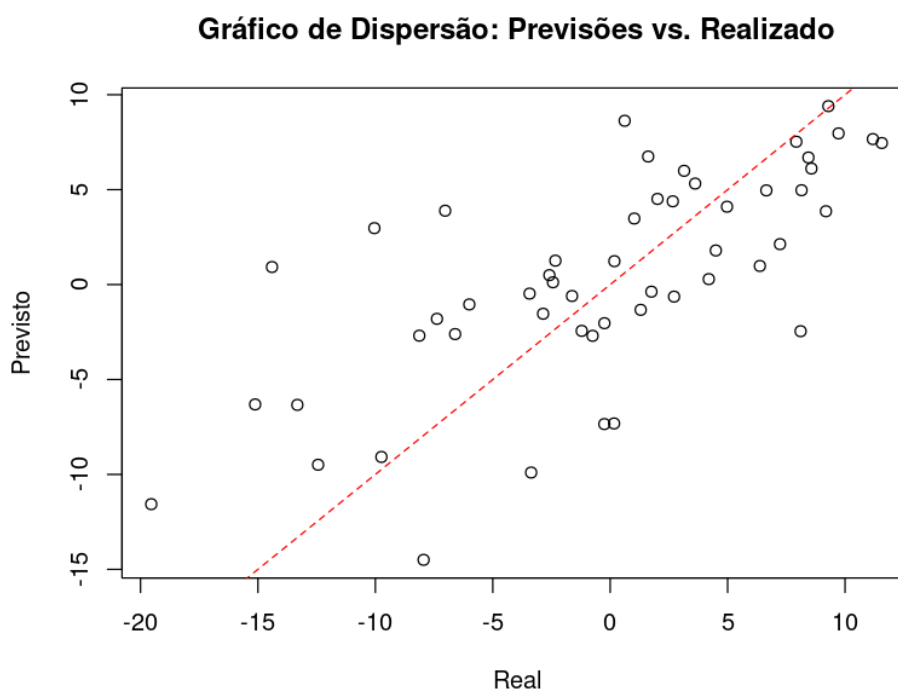


Figura 5: Gráfico de dispersão entre os valores previstos e reais - com linha de referência

A partir disso, pode-se obter, através da função `xgb.plot.importance` do pacote `xgboost`, a importância de cada variável para a predição do *score*. A Figura 6 mostra as 10 principais variáveis com mais peso para a predição. Quatro métricas das cinco mais importantes são relativas ao treino imediatamente anterior ao jogo e isso se estende as 10 principais, sendo 6 das 10 referentes ao treino um dia antes do jogo.

Além disso, 5 das 10 principais métricas mais importantes pertencem à categoria de velocidade, sendo elas: `Accelerations_Zone2_t-1`, `Accelerations_Impulse_t-1`, `Accelerations_Impulse_t-2`, `Decelerations_Zone1_t-1` e `Speed_Intensity_Zone5_t-3`. Isso indica que exercícios relativos a velocidade tem uma alta importância na predição do *score* da partida.

Variáveis de impacto também se mostram relevantes para o *score*, dado que a mais importante é `Impact_Zone2_t-1` e ainda, existem mais duas métricas desse tipo no ranking:

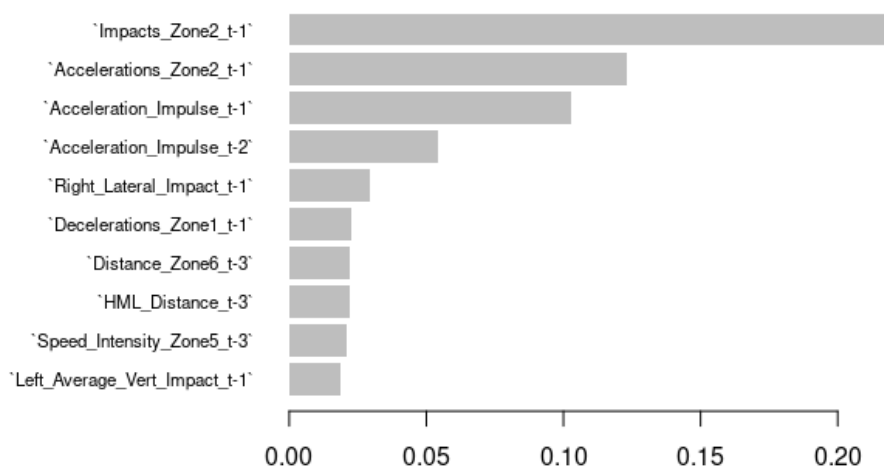


Figura 6: Gráfico de barras das 10 variáveis mais importantes para o modelo

Right_Lateral_Impact.t-1 e Left_Average_Vert_Impact.t-1. É importante ressaltar que as 3 acontecem no treino imediatamente anterior ao jogo.

As duas últimas variáveis consistem em distâncias cobertas. São elas:

- Distance_Zone6_t-3: Distância coberta na zona de velocidade 6.
- HML_Distance.t-3: Distância coberta realizando qualquer atividade acima de 25,5 W/Kg.

Diferente das anteriores, essas são as únicas realizadas no treino 3 dias antes do jogo.

Ou seja, variáveis da categoria velocidade, em sua maioria, têm um peso maior no treino -1, sendo essa categoria a que reúne a maior importância dentre as 10 principais. Métricas de impacto aparecem exclusivamente no treino -1. As únicas que indicam maior peso no treino -3 são as relativas a distância coberta.

4 Conclusão

O seguinte estudo procurou relacionar métricas físicas de treino dos jogadores de um time profissional de futebol brasileiro com as mesmas variáveis nas partidas oficiais. A base de dados obtida para a realização do estudo apresentava diversas inconsistências que necessitaram de um tratamento para não influenciar no resultado final. Vale ressaltar que em função da complexidade de certos aspectos específicos da área, a abordagem utilizada para tratar essas inconsistências poderia ser ainda mais refinada. A inclusão do conhecimento de um profissional de educação física poderia ter enriquecido a análise dos dados e possibilitado um tratamento ainda mais detalhado. Dessa forma, ao reconhecer a relevância de uma visão especializada, o estudo abriu caminho para considerações futuras e potenciais aprimoramentos metodológicos ao lado de um especialista da área de treinamento físico.

A utilização do PCA para resumir as variáveis de jogos a fim de transformá-las em uma métrica única de avaliação física foi uma solução que mostrou-se muito valiosa. No entanto, utilizar apenas a componente principal 1 sendo que essa explica 48% das variações, é um ponto que poderia ser otimizado com um aprofundamento maior na própria base, utilizando mais componentes ou, até mesmo, alguma outra técnica que incluísse a maior variação explicada possível.

Sobre a implementação do modelo de aprendizado de máquinas, o XGBoost (xgb) mostrou um resultado suficiente para o Coeficiente de Determinação, próximo a 49% e foi o escolhido. Ainda, em Rossi et al. (2018) também foi utilizado um modelo de árvores de decisão, o que reforçou a ideia de permanecer com o xgb. Esse algoritmo também poderia ser otimizado com um estudo mais aprofundado sobre os hiperparâmetros ou tentando utilizar mais algoritmos de aprendizado de máquinas, dado a infinidade de modelos de regressão disponíveis; porém, o maior dificultador desses resultados foi o tratamento realizado na base de dados, uma vez que a não compreensão de 100% das métricas e dinâmicas de treinos pode ter acarretado em decisões equivocadas sobre o tratamento das informações.

Como já citado anteriormente, contar com a orientação de um profissional de educação física especializado em treinos de futebol teria sido uma contribuição relevante para aprimorar ainda mais a análise, conferindo maior confiabilidade aos resultados obtidos.

E por fim, é importante ressaltar que as conclusões do estudo apresentam um valor significativo. Encontrou-se nas variáveis mais importantes para a predição do *score*, métricas associadas ao treino imediatamente anterior ao jogo. Isso implica em uma reflexão acerca dos treinos de aquecimento no dia do respectivo jogo removidos da análise. Avaliar esses exercícios pode trazer mais conclusões no que se refere ao *score* da partida. As métricas de velocidade também se mostram muito relevantes para a predição, todavia a métrica que obteve maior importância, das mais de 100 para cada treino anterior utilizadas na modelagem, foi a que representa o número de impactos na zona de 5-7G de magnitude no treino um dia antes da partida. Isso sugere que exercícios físicos que auxiliam no aprimoramento dessa métrica implicam em um *score* de jogo mais robusto. Contudo, esse resultado não implica em simplesmente submeter os atletas a situações de alta velocidade, aceleração ou impacto, visto que Malone et al. (2017) mostrou que sujeitar atletas a mudanças rápidas de incremento nas distâncias percorridas em alta velocidade (*high speed running*) e corridas com arrancada (*sprint running*), por exemplo, acabam aumentando o risco de lesões. Portanto, é sempre importante ter um profissional da área que poderá direcionar de forma mais assertiva o uso das informações contidas nesse estudo aos exercícios.

Referências

Bernardo Coutinho. *Modelos de Predição — SVM*. 2019. Disponível em: <https://medium.com/turing-talks/turing-talks-12-classifica%C3%A7%C3%A3o-por-svm-f4598094a3f1>. Acesso em: 13 de julho 2023.

Carlos Alberto Alves Varella. *Análise Multivariada Aplicada as Ciências Agrárias - Análise de Componentes Principais*. 2008. Disponível em: <http://www.ufrrj.br/institutos/it/deng/varella/Downloads/multivariada%20aplicada%20as%20ciencias%20agrarias/Aulas/analise%20de%20componentes%20principais.pdf>. Acesso em: 13 de julho 2023.

Clébio de Oliveira Júnior. *Métricas para Regressão: Entendendo as métricas R^2 , MAE, MAPE, MSE e RMSE*. 2021. Disponível em: <https://medium.com/data-hackers/prevendo-n%C3%BAmeros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>. Acesso em: 13 de julho 2023.

Didática Tech. *Como funciona o algoritmo Árvore de Decisão*. 2022. Disponível em: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>. Acesso em: 13 de julho 2023.

Flavio Souza. *Principais clubes do Brasil arrecadaram mais de R\$ 4.5 bilhões em 2020*. 2021. Disponível em: <https://www.torcedores.com/noticias/2021/09/clubes-receitas-brasil-2020>. Acesso em: 10 de maio 2022.

João Guilherme Berti Sczip. *Pressupostos da Regressão Linear*. 2019. Disponível em: <https://joaogbsczip.medium.com/pressupostos-da-regress%C3%A3o-linear-d07f784fe11a>. Acesso em: 13 de julho 2023.

MALONE, S. et al. High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? *Journal of Science and Medicine in Sport*, v. 21, 05 2017.

Onédio Junior. *O Guia do XGBoost com Python*. 2022. Disponível em: <https://dadosaocubo.com/o-guia-do-xgboost-com-python/>. Acesso em: 13 de julho 2023.

RAGO, V. et al. Methods to collect and interpret external training load using microtechnology incorporating gps in professional football: a systematic review. *Research in Sports Medicine*, 11 2019.

RAGO, V. et al. Training load and submaximal heart rate testing throughout a competitive period in a top-level male football team. *Journal of Sports Sciences*, Routledge, v. 38, n. 11-12, p. 1408–1415, 2020. PMID: 31130068. Disponível em: <https://doi.org/10.1080/02640414.2019.1618534>.

ROSSI, A. et al. Effective injury forecasting in soccer with gps training data and machine learning. *PLOS ONE*, Public Library of Science, v. 13, n. 7, p. 1–15, 07 2018. Disponível em: <https://doi.org/10.1371/journal.pone.0201264>.

Unisport Brasil. *Monitoramento de atletas: como fazer durante os treinamentos?* 2020. Disponível em: <https://www.unisportbrasil.com.br/monitoramento-de-atletas/>. Acesso em: 13 de julho 2023.

APÊNDICE 1 – Descrição das variáveis

INDEX	MÉTRICA	DESCRIÇÃO
1	% TIME IN RED ZONE	%TEMPO GASTO ACIMA DA ZONA 5 DE FREQUÊNCIA CARDÍACA
2	ACCELERATION IMPULSE	INTENSIDADE DE ACELERAÇÃO DO EXERCÍCIO OU SESSÃO SELECIONADA - CALCULADA USANDO A MAGNITUDE CUMULATIVA DIVIDIDA PELO NÚMERO DE OCORRÊNCIAS
3	ACCELERATIONS	NÚMERO DE ACELERAÇÃO DAS ZONAS ATIVAS EM PREFERÊNCIAS
4	ACCELERATIONS ZONE1	ACELERAÇÕES ENTRE 0.5-1.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
5	ACCELERATIONS ZONE2	ACELERAÇÕES ENTRE 1.0-2.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
6	ACCELERATIONS ZONE3	ACELERAÇÕES ENTRE 2.0-3.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
7	ACCELERATIONS ZONE4	ACELERAÇÕES ENTRE 3.0-4.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
8	ACCELERATIONS ZONE5	ACELERAÇÕES ENTRE 4.0-5.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
9	ACCELERATIONS ZONE6	ACELERAÇÕES ENTRE 5.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
10	ACCELERATIONS Z3 TO Z6	ACELERAÇÕES ENTRE 2.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
11	ACCELERATIONS Z4 TO Z6	ACELERAÇÕES ENTRE 3.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
12	ACCELERATIONS Z5 TO Z6	ACELERAÇÕES ENTRE 4.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
13	ACCELERATIONS TOTAL DISTANCE Z1	DISTÂNCIA COBERTA ENQUANTO ACELERA EM VELOCIDADES DE 0-1.5 M/S
14	ACCELERATIONS TOTAL DISTANCE Z2	DISTÂNCIA COBERTA ACELERANDO A VELOCIDADES DE 1.5-3M/S
15	ACCELERATIONS TOTAL DISTANCE Z3	DISTÂNCIA COBERTA ACELERANDO A VELOCIDADES DE 3-4M/S
16	ACCELERATIONS TOTAL DISTANCE Z4	DISTÂNCIA COBERTA ACELERANDO A VELOCIDADES DE 4-5,5M/S
17	ACCELERATIONS TOTAL DISTANCE Z5	DISTÂNCIA COBERTA ACELERANDO A VELOCIDADES DE 5.5-7M/S
18	ACCELERATIONS TOTAL DISTANCE Z6	DISTÂNCIA COBERTA ACELERANDO A VELOCIDADES DE 7-10M/S
19	ACOES INTENSAS	QUANTIDADE DE AÇÕES INTENSAS
20	ACUTE VS CHRONIC RATIO	RELAÇÃO DE TRABALHO FEITO EM UM CURTO PERÍODO DE TEMPO EM COMPARAÇÃO AO GRANDE PRAZO USANDO QUALQUER MÉTRICA
21	AVERAGE HEART RATE	FREQUÊNCIA CARDÍACA MÉDIA DO JOGADOR DURANTE A SESSÃO/EXERCÍCIO
22	AVERAGE METABOLIC POWER	SAÍDA DE POTÊNCIA MÉDIA EM WATTS POR KG DE PESO CORPORAL POR SEGUNDO DURANTE TREINO/SESSÃO
23	AVERAGE SPEED	VELOCIDADE MÉDIA NA PREFERÊNCIA DE VELOCIDADE SELECIONADA DURANTE EXERCÍCIO/SESSÃO
24	BODY	OPÇÃO NA AVALIAÇÃO SUBJETIVA RTT - AVALIAR O ESTADO DE BEM-ESTAR DE UM INDIVÍDUO ANTES DO TREINAMENTO/JOGO
25	CUSTOM METRIC 10	MÉTRICA PERSONALIZADA PELA EQUIPE

INDEX	MÉTRICA	DESCRIÇÃO
26	CUSTOM METRIC 5	MÉTRICA PERSONALIZADA PELA EQUIPE
27	CUSTOM METRIC 6	MÉTRICA PERSONALIZADA PELA EQUIPE
28	CUSTOM METRIC 7	MÉTRICA PERSONALIZADA PELA EQUIPE
29	CUSTOM METRIC 8	MÉTRICA PERSONALIZADA PELA EQUIPE
30	CUSTOM METRIC 9	MÉTRICA PERSONALIZADA PELA EQUIPE
31	CUSTOM PLAYER ID	MÉTRICA PERSONALIZADA PELA EQUIPE
32	DECELERATIONS	NÚMERO DE DESACELERAÇÃO DAS ZONAS ATIVAS EM preferências
33	DECELERATIONS ZONE 1	DESACELERAÇÕES ENTRE 0.5-1.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
34	DECELERATIONS ZONE 2	DESACELERAÇÕES ENTRE 1.0-2.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
35	DECELERATIONS ZONE 3	DESACELERAÇÕES ENTRE 2.0-3.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
36	DECELERATIONS ZONE 4	DESACELERAÇÕES ENTRE 3.0-4.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
37	DECELERATIONS ZONE 5	DESACELERAÇÕES ENTRE 4.0-5.0 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
38	DECELERATIONS ZONE 6	DESACELERAÇÕES ENTRE 5.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
39	DECELERATIONS Z3 TO Z6	DESACELERAÇÕES ENTRE 2.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
40	DECELERATIONS Z4 TO Z6	DESACELERAÇÕES ENTRE 3.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
41	DECELERATIONS Z5 TO Z6	DESACELERAÇÕES ENTRE 4.0-10 M/S/S COM DURAÇÃO MÍNIMA DE 0.5 S POR PADRÃO
42	DECELERATIONS TOTAL DISTANCE Z1	DISTÂNCIA COBERTA ENQUANTO DESACELERA EM VELOCIDADES DE 0-1.5 M/S
43	DECELERATIONS TOTAL DISTANCE Z2	DISTÂNCIA COBERTA DESACELERANDO A VELOCIDADES DE 1.5-3M/S
44	DECELERATIONS TOTAL DISTANCE Z3	DISTÂNCIA COBERTA DESACELERANDO A VELOCIDADES DE 3-4M/S
45	DECELERATIONS TOTAL DISTANCE Z4	DISTÂNCIA COBERTA DESACELERANDO A VELOCIDADES DE 4-5.5M/S
46	DECELERATIONS TOTAL DISTANCE Z5	DISTÂNCIA COBERTA DESACELERANDO A VELOCIDADES DE 5.5-7M/S
47	DECELERATIONS TOTAL DISTANCE Z6	DISTÂNCIA COBERTA DESACELERANDO A VELOCIDADES DE 7-10M/S
48	DISTANCE PER MIN	DISTÂNCIA TOTAL COBERTA DIVIDIDA PELO TEMPO TOTAL DA SESSÃO/EXERCÍCIO SELECIONADA
49	DISTANCE TOTAL	DISTÂNCIA TOTAL COBERTA PELO(S) JOGADOR(ES) NA SESSÃO/EXERCÍCIO SELECIONADO
50	DISTANCE Z2TOZ6	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE RELATIVA 2-6
51	DISTANCE Z2TOZ6AB	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE ABSOLUTA 2-6
52	DISTANCE Z3TOZ6	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE RELATIVA 3-6
53	DISTANCE Z3TOZ6AB	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE ABSOLUTA 3-6
54	DISTANCE Z4TOZ6	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE RELATIVA 4-6
55	DISTANCE Z4TOZ6AB	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE ABSOLUTA 4-6
56	DISTANCE Z5TOZ6	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE RELATIVA 5-6
57	DISTANCE Z5TOZ6AB	DISTÂNCIA COBERTA EM ZONAS DE VELOCIDADE ABSOLUTA 5-6
58	DISTANCE ZONAL	DISTÂNCIA TOTAL COBERTA EM CADA UMA DAS ZONAS DE VELOCIDADE
59	DISTANCE ZONE 1 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 1
60	DISTANCE ZONE 2 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 2
61	DISTANCE ZONE 3 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 3
62	DISTANCE ZONE 4 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 4
63	DISTANCE ZONE 5 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 5
64	DISTANCE ZONE 6 (ABSOLUTE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 6
65	DISTANCE ZONE 1 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 1
66	DISTANCE ZONE 2 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 2
67	DISTANCE ZONE 3 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 3
68	DISTANCE ZONE 4 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 4
69	DISTANCE ZONE 5 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 5

INDEX	MÉTRICA	DESCRIÇÃO
70	DISTANCE ZONE 6 (RELATIVE)	DISTÂNCIA COBERTA NA ZONA DE VELOCIDADE 6
71	DRILL DATE	DATA DE REALIZAÇÃO DO EXERCÍCIO
72	DRILL END TIME	TEMPO EM QUE O EXERCÍCIO FOI FINALIZADO
73	DRILL ID	ID DO EXERCÍCIO
74	DRILL START TIME	TEMPO EM QUE O EXERCÍCIO FOI INICIADO
75	DRILL TITLE	TÍTULO DO EXERCÍCIO
76	DURATION OF HIGH INTENSITY BURSTS	DURAÇÃO (HH:MM:SS) DE ESTOUROS DE ALTA INTENSIDADE (VER ABAIXO)
77	DYNAMIC STRESS LOAD	PESO TOTAL DE IMPACTOS ACIMA DE 2G COM BASE EM CLASSIFICAÇÕES DE FORÇA G CURVA CONVEXA
78	DYNAMIC STRESS LOAD Z3 TO Z6	DESCRIÇÃO NÃO FORNECIDA
79	DYNAMIC STRESS LOAD Z4 TO Z6	DESCRIÇÃO NÃO FORNECIDA
80	DYNAMIC STRESS LOAD Z5 TO Z6	DESCRIÇÃO NÃO FORNECIDA
81	DYNAMIC STRESS LOAD ZONE1	DESCRIÇÃO NÃO FORNECIDA
82	DYNAMIC STRESS LOAD ZONE2	DESCRIÇÃO NÃO FORNECIDA
83	DYNAMIC STRESS LOAD ZONE3	DESCRIÇÃO NÃO FORNECIDA
84	DYNAMIC STRESS LOAD ZONE4	DESCRIÇÃO NÃO FORNECIDA
85	DYNAMIC STRESS LOAD ZONE5	DESCRIÇÃO NÃO FORNECIDA
86	DYNAMIC STRESS LOAD ZONE6	DESCRIÇÃO NÃO FORNECIDA
87	ENERGY EXPENDITURE (KCAL)	ENERGIA GASTA PARA ESFORÇO LOCOMOTOR EXPRESSO COMO (KCAL)
88	ENTRIES ZONE 3 (ABSOLUTE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE 3 LIMITE
89	ENTRIES ZONE 4 (ABSOLUTE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE 4 LIMITE
90	ENTRIES ZONE 5 (ABSOLUTE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE 5 LIMITE
91	ENTRIES ZONE 6 (ABSOLUTE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE 6 LIMITES
92	ENTRIES ZONE 3 (RELATIVE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE RELATIVA 3 LIMITES
93	ENTRIES ZONE 4 (RELATIVE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE RELATIVA 4 LIMITES
94	ENTRIES ZONE 5 (RELATIVE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE RELATIVA 5 LIMITES
95	ENTRIES ZONE 6 (RELATIVE)	NÚMERO DE ENTRADAS ENTRE ZONA DE VELOCIDADE RELATIVA 6 LIMITES
96	EQUIVALENT METABOLIC DISTANCE	UTILIZANDO A ENERGIA TOTAL DESGASTADA, A DISTÂNCIA EM METROS QUE UM ATLETA PRECISARIA PERCORRER EM VELOCIDADE CONSTANTE PARA GASTAR A MESMA QUANTIDADE DE ENERGIA
97	EXPLOSIVE DISTANCE	HMLD MENOS HSR. SE O LIMITE HSR FOR DEIXADO EM 5.5 M/S, ESTA DISTÂNCIA SERIA DEVIDO À ACELERAÇÃO E DESACELERAÇÃO ACIMA DO LIMITE DE 25.5 W/KG
98	FATIGUE INDEX	PONTUAÇÃO DE INTENSIDADE DE VELOCIDADE DIVIDIDA POR CARGA DE TENSÃO DINÂMICA
99	HEART RATE EXERTION	ESCALA REGISTRADA DE % DE FREQUÊNCIA CARDÍACA SEGUINDO A CURVA CONVEXA. CADA PONTUAÇÃO DE FREQUÊNCIA CARDÍACA OBTENDO UMA PESAGEM ESPECÍFICA E ACRESCENTADA NA SESSÃO
100	HEART RATE RECOVERY BEATS	BATIMENTOS DE RECUPERAÇÃO DA FREQUÊNCIA CARDÍACA
101	HEART RATE RECOVERY PERCENT	PORCENTAGEM DE RECUPERAÇÃO DA FREQUÊNCIA CARDÍACA
102	HEART RATE VARIABILITY	VARIAÇÃO NO INTERVALO DE BATIDA-A -BATIDA ENTRE BATIMENTOS CARDÍACOS INDIVIDUAIS EM MÉDIA DE 2 MINUTOS.
103	HIGH SPEED RUNNING (ABSOLUTE)	DISTÂNCIA COBERTA ULTRAPASSANDO O LIMITE DA ZONA 5
104	HIGH SPEED RUNNING (RELATIVE)	DISTÂNCIA COBERTA ULTRAPASSANDO O LIMITE DA ZONA 5 (VELOCIDADE VARIÁVEL)
105	HML DISTANCE	DISTÂNCIA COBERTA REALIZANDO QUALQUER ATIVIDADE ACIMA DE 25.5W/KG. LIMITE CONFIGURÁVEL NO PERFIL DO JOGADOR.
106	HML DISTANCE PER MINUTE	DISTÂNCIA HML POR MINUTO DE SESSÃO/TREINO SELECIONADO
107	HML EFFORTS	NÚMERO DE ESFORÇOS REALIZADOS ACIMA DO LIMITE DE 25.5 W/KG

INDEX	MÉTRICA	DESCRIÇÃO
108	HML TIME	TEMPO GASTO REALIZANDO ESFORÇOS ACIMA DO LIMITE DE 25.5 W/KG
109	HSR PER MINUTE (ABSOLUTE)	DISTÂNCIA COBERTA ACIMA DO LIMITE DE VELOCIDADE DA ZONA 5 POR MINUTO DE SESSÃO/TREINO SELECIONADO
110	HSR PER MINUTE (RELATIVE)	DISTÂNCIA COBERTA ACIMA DA ZONA DE VELOCIDADE 5 POR MINUTO DE SESSÃO/TREINO SELECIONADO
111	IMPACTS	A SOMA DA MAGNITUDE DOS VALORES DO ACELERÔMETRO DE 3 EIXOS EM UM PERÍODO DE 1 SEGUNDO, SUPERIOR A 2G. DEPENDENTE DAS ZONAS ATIVAS.
112	IMPACTS ZONE1	NÚMERO DE IMPACTOS DE 3-5G MAGNITUDE
113	IMPACTS ZONE2	NÚMERO DE IMPACTOS DE 5-7G MAGNITUDE
114	IMPACTS ZONE3	NÚMERO DE IMPACTOS DE 7-9G MAGNITUDE
115	IMPACTS ZONE4	NÚMERO DE IMPACTOS DE 9-11G MAGNITUDE
116	IMPACTS ZONE5	NÚMERO DE IMPACTOS DE 11-13G MAGNITUDE
117	IMPACTS ZONE6	NÚMERO DE IMPACTOS DE 13-15G MAGNITUDE
118	IMPACTS Z3 TO Z6	NÚMERO DE IMPACTOS DE 7-15G MAGNITUDE POR PADRÃO (CONFIGURÁVEL)
119	IMPACTS Z4 TO Z6	NÚMERO DE IMPACTOS DE 9-15G MAGNITUDE POR PADRÃO (CONFIGURÁVEL)
120	IMPACTS Z5 TO Z6	NÚMERO DE IMPACTOS DE 11-15G MAGNITUDE POR PADRÃO (CONFIGURÁVEL)
121	INTERVAL	DESCRIÇÃO NÃO FORNECIDA
122	LEFT ANT POST IMPACT	DESCRIÇÃO NÃO FORNECIDA
123	LEFT AVERAGE VERT IMPACT	DESCRIÇÃO NÃO FORNECIDA
124	LEFT LATERAL IMPACT	DESCRIÇÃO NÃO FORNECIDA
125	LEFT MAG IMPACT	DESCRIÇÃO NÃO FORNECIDA
126	LEFT VERTICAL IMPACT	DESCRIÇÃO NÃO FORNECIDA
127	LOWER SPEED LOADING	TODA A CARGA DO ACELERÔMETRO REGISTRADA AO VIAJAR EM VELOCIDADES INFERIORES A 2M/S
128	MAX ACCELERATION	MAGNITUDE DE ACELERAÇÃO MÁXIMA (CALCULADA POR EXERCÍCIO/SESSÃO)
129	MAX DECELERATION	MAGNITUDE MÁXIMA DE DESACELERAÇÃO (CALCULADA POR EXERCÍCIO/SESSÃO)
130	MAX SPEED	VELOCIDADE MÁXIMA ATINGIDA NA SESSÃO/EXERCÍCIO SELECIONADO EXPRESSA NA PREFERÊNCIA DE VELOCIDADE (M/S, KM/H, MPH)
131	MAX HEART RATE	FREQUÊNCIA CARDÍACA MÁXIMA ATINGIDA NA SESSÃO/EXERCÍCIO SELECIONADO, EXPRESSA EM BATIMENTOS POR MINUTO (BPM)
132	METABOLIC DISTANCE ZONAL	DISTÂNCIA COBERTA EM CADA UMA DAS ZONAS METABÓLICAS
133	METABOLIC DISTANCE ZONE 1	DISTÂNCIA COBERTA ENQUANTO GASTA 0-5 WATTS POR KG DE MASSA CORPORAL
134	METABOLIC DISTANCE ZONE 2	DISTÂNCIA COBERTA ENQUANTO GASTA 5-10 WATTS POR KG DE MASSA CORPORAL
135	METABOLIC DISTANCE ZONE 3	DISTÂNCIA COBERTA ENQUANTO GASTA 10-15 WATTS POR KG DE MASSA CORPORAL
136	METABOLIC DISTANCE ZONE 4	DISTÂNCIA COBERTA ENQUANTO GASTA 15-25 WATTS POR KG DE MASSA CORPORAL
137	METABOLIC DISTANCE ZONE 5	DISTÂNCIA COBERTA ENQUANTO GASTA 25-50 WATTS POR KG DE MASSA CORPORAL
138	METABOLIC DISTANCE ZONE 6	DISTÂNCIA COBERTA ENQUANTO GASTA 50-500 WATTS POR KG DE MASSA CORPORAL
139	METABOLIC TIME ZONAL	TEMPO COBERTO EM CADA UMA DAS ZONAS METABÓLICAS
140	METABOLIC TIME ZONE 1	TEMPO GASTO 0-5 WATTS POR KG DE MASSA CORPORAL
141	METABOLIC TIME ZONE 2	TEMPO GASTO 5-10 WATTS POR KG DE MASSA CORPORAL
142	METABOLIC TIME ZONE 3	TEMPO GASTO DE 10-15 WATTS POR KG DE MASSA CORPORAL
143	METABOLIC TIME ZONE 4	TEMPO GASTO 15-25 WATTS POR KG DE MASSA CORPORAL
144	METABOLIC TIME ZONE 5	TEMPO GASTO 25-50 WATTS POR KG DE MASSA CORPORAL
145	METABOLIC TIME ZONE 6	TEMPO GASTO 50-500 WATTS POR KG DE MASSA CORPORAL

INDEX	MÉTRICA	DESCRIÇÃO
146	NUMBER OF HIGH INTENSITY BURSTS	O NÚMERO TOTAL DE VEZES EM QUE OCORRERAM UM MÍNIMO DE 3 ATIVIDADES DE ALTA INTENSIDADE EM UM PERÍODO DE TEMPO CONFIGURÁVEL.
147	NUTRITION	OPÇÃO DENTRO DO APLICATIVO RTT SUBJETIVO - UTILIZADO PARA AVALIAR A QUALIDADE DE NUTRIÇÃO DE UM INDIVÍDUO
148	OSMOLARITY	A MEDIDA DA CONCENTRAÇÃO DE SOLUTO POR UNIDADE DE VOLUME
149	PERCEIVED EXERTION	PERGUNTA SUBJETIVA UTILIZADA PARA AVALIAR A PERCEPÇÃO DE UM INDIVÍDUO SOBRE A RESISTÊNCIA DE UMA SESSÃO
150	PLAYER DATE OF BIRTH	DATA DE ANIVERSÁRIO DO JOGADOR
151	PLAYER DISPLAY NAME	NOME APRESENTADO NA PRIMEIRA CAIXA DE OPÇÕES DO PERFIL DOS JOGADORES.
152	PLAYER FIRST NAME	PRIMEIRO NOME DO JOGADOR
153	PLAYER HEIGHT	MEDIDA EM PÉS - CONFIGURÁVEL EM EDIT → PREFERÊNCIAS
154	PLAYER ID	ID DO JOGADOR
155	PLAYER LAST NAME	SOBRENOME DO JOGADOR
156	PLAYER WEIGHT	MEDIDAS EM KG - CONFIGURÁVEIS EM EDIT → PREFERÊNCIAS
157	PLAYER MAX HEART RATE [FROM PROFILE]	FREQUÊNCIA CARDÍACA MÁXIMA DEFINIDA NO PERFIL DO JOGADOR
158	PLAYER NAME	NOME DO JOGADOR
159	PLAYER POSITION	POSIÇÃO DE JOGO - USADA PRINCIPALMENTE EM ESPORTES DE EQUIPE
160	READINESS	OPÇÃO NA AVALIAÇÃO SUBJETIVA RTT- AVALIAR O ESTADO DE BEM-ESTAR DE UM INDIVÍDUO ANTES DO TREINO/JOGO
161	RIGHT ANT POST IMPACT	DESCRIÇÃO NÃO FORNECIDA
162	RIGHT AVERAGE VERT IMPACT	DESCRIÇÃO NÃO FORNECIDA
163	RIGHT LATERAL IMPACT	DESCRIÇÃO NÃO FORNECIDA
164	RIGHT MAG IMPACT	DESCRIÇÃO NÃO FORNECIDA
165	RIGHT VERTICAL IMPACT	DESCRIÇÃO NÃO FORNECIDA
166	SESSION DATE	DATA DA REALIZAÇÃO DA SESSÃO
167	SESSION DAY OF WEEK	DIA DA SEMANA EM QUE A SESSÃO FOI REALIZADA
168	SESSION END TIME	TEMPO EM QUE A SESSÃO FOI FINALIZADA
169	SESSION ID	ID DA SESSÃO
170	SESSION START TIME	TEMPO EM QUE A SESSÃO FOI INICIADA
171	SESSION TITLE	TÍTULO DA SESSÃO
172	SESSION TYPE	PERFIL OU STATUS DE UM DADO DIA. EX: DIA DA PARTIDA OU DIA DA PARTIDA -1
173	SESSION WEEK NUMBER	NÚMERO DA SEMANA EM QUE FOI REALIZADA A SESSÃO
174	SLEEP	AVALIAÇÃO DA QUALIDADE DO SONO DO ATLETA - ESCALA NÃO FORNECIDA
175	SPEED INTENSITY	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NA CURVA CONVEXA E PESOS PARA CADA VELOCIDADE INDIVIDUAL. QUANTO MAIS TEMPO VOCÊ PASSA EM ALTA VELOCIDADE, MAIOR INTENSIDADE DE VELOCIDADE VOCÊ TERÁ.
176	SPEED INTENSITY Z3 TO Z6	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NAS ZONAS DE VELOCIDADE 3 - 6 SOMENTE
177	SPEED INTENSITY Z4 TO Z6	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NAS ZONAS DE VELOCIDADE 4 - 6 SOMENTE
178	SPEED INTENSITY Z5 TO Z6	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NAS ZONAS DE VELOCIDADE 5 - 6 SOMENTE
179	SPEED INTENSITY ZONE1	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 1 SOMENTE (0 - 1.5 M/S)
180	SPEED INTENSITY ZONE2	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 2 SOMENTE (1.5 - 3M/S)
181	SPEED INTENSITY ZONE3	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 3 SOMENTE (3 - 4M/S)
182	SPEED INTENSITY ZONE4	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 4 SOMENTE (4 - 5.5 M/S)

INDEX	MÉTRICA	DESCRIÇÃO
183	SPEED INTENSITY ZONE5	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 5 SOMENTE (5.5 - 7M/S)
184	SPEED INTENSITY ZONE6	MEDIÇÃO DO ESFORÇO TOTAL COM BASE NO TEMPO GASTO NA ZONA DE VELOCIDADE 6 SOMENTE (7 - 11M/S)
185	SPRINT DISTANCE	A DISTÂNCIA É COBERTA A UMA VELOCIDADE DE $\geq 5,5$ M/S. A DURAÇÃO MÍNIMA DE UM SPRINT É DE 1 SEGUNDO. AMBOS OS DETERMINANTES SÃO CONFIGURÁVEIS EM 'EDIT → PREFERÊNCIAS' NO SOFTWARE APEX. A DISTÂNCIA ACIMA DO LIMITE DE SPRINT É REALIZADA POR 1 SEG OU MAIS E DEPOIS TERMINA QUANDO VOCÊ CAIR ABAIXO DE 80% DO LIMITE.
186	SPRINTS	O NÚMERO TOTAL DE SPRINTS REALIZADOS POR UM INDIVÍDUO. SPRINTS SÃO ALCANÇADOS CORRENDO A PARTIR DE 5.5M/S POR UMA DURAÇÃO MÍNIMA DE 1 SEGUNDO.
187	STEP BALANCE	MÉDIA DOS IMPACTOS ESQUERDO E DIREITO
188	STRESS	OPÇÃO DENTRO DO APLICATIVO RTT SUBJETIVO - UTILIZADO PARA AVALIAR O NÍVEL DE ESTRESSE DE UM INDIVÍDUO - ESTRESSOR EXTERNO V INTERNO
189	SUBJECTIVE Q6	DESCRIÇÃO NÃO FORNECIDA
190	TIME IN HEART RATE Z3 TO Z6	TOTAL DE TEMPO GASTO EM ZONAS DE FREQUÊNCIA CARDÍACA 3 A 6 - EX: NO PADRÃO - ENTRE 117 BPM A 216 BPM
191	TIME IN HEART RATE Z4 TO Z6	TOTAL DE TEMPO GASTO EM ZONAS DE FREQUÊNCIA CARDÍACA 4 A 6 - EX: NO PADRÃO - ENTRE 135BPM A 216BPM
192	TIME IN HEART RATE ZONE1	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 1 - EX: NO PADRÃO - ENTRE 0 - 99 BPM
193	TIME IN HEART RATE ZONE2	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 2 - EX: NO PADRÃO - ENTRE 99 - 117 BPM
194	TIME IN HEART RATE ZONE3	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 3 - EX: NO PADRÃO - ENTRE 117 - 135 BPM
195	TIME IN HEART RATE ZONE4	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 4 - EX: NO PADRÃO - ENTRE 135 - 153 BPM
196	TIME IN HEART RATE ZONE5	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 5 - EX: NO PADRÃO - ENTRE 153 - 171 BPM
197	TIME IN HEART RATE ZONE6	TOTAL DE TEMPO GASTO NA ZONA DE FREQUÊNCIA CARDÍACA 6 - EX: NO PADRÃO - ENTRE 171 - 216 BPM
198	TIME IN RED ZONE	TEMPO TOTAL GASTO ACIMA DO LIMITE DA ZONA 5 DE UMA FREQUÊNCIA CARDÍACA MÁXIMA DE INDIVÍDUOS - EX: NO PADRÃO DE 220 BPM = TEMPO GASTO ACIMA DE 187 BPM
199	TIME ZONE 1 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 1 (0 - 1.5 M/S)
200	TIME ZONE 1 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 1 (0 - 1.5 M/S)
201	TIME ZONE 2 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 2 (1.5 - 3M/S)
202	TIME ZONE 2 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 2 (1.5 - 3M/S)
203	TIME ZONE 3 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 3 (3 - 4M/S)
204	TIME ZONE 3 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 3 (3 - 4M/S)
205	TIME ZONE 4 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 4 (4 - 5.5 M/S)
206	TIME ZONE 4 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 4 (4 - 5.5 M/S)
207	TIME ZONE 5 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 5 (5.5 - 7M/S)
208	TIME ZONE 5 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 5 (5.5 - 7M/S)
209	TIME ZONE 6 (ABSOLUTE)	TEMPO GASTO NA ZONA DE VELOCIDADE 6 (7 - 11M/S)
210	TIME ZONE 6 (RELATIVE)	TEMPO GASTO NA ZONA DE VELOCIDADE 6 (7 - 11M/S)
211	TOTAL DECELERATION LOADING	MÉTRICA DE CARGA DE TENSÃO EXCÊNTRICA - CALCULADA USANDO CARGA DE TENSÃO DINÂMICA
212	TOTAL LEFT STEPS	NÚMERO TOTAL DE PASSOS À ESQUERDA
213	TOTAL RIGHT STEPS	NÚMERO TOTAL DE PASSOS À DIREITA

INDEX	MÉTRICA	DESCRIÇÃO
214	TOTAL LOADING	USANDO APENAS OS DADOS DO ACELERÔMETRO, É O TOTAL DAS FORÇAS NO JOGADOR EM TODA A SESSÃO SEM QUALQUER PESAGEM APLICADA
215	TOTAL TIME	TEMPO TOTAL DA SESSÃO