

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-graduação em Estatística

Ronaldo Angelo Dias da Silva

***MACHINE LEARNING PARA A PREDIÇÃO DA CARGA DE TREINAMENTO DE
UM ATLETA DA SELEÇÃO BRASILEIRA DE TAEKWONDO***

Belo Horizonte
2023

Ronaldo Angelo Dias da Silva

***MACHINE LEARNING PARA A PREDIÇÃO DA CARGA DE TREINAMENTO DE
UM ATLETA DA SELEÇÃO BRASILEIRA DE TAEKWONDO***

Monografia apresentada ao curso de Especialização em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística Aplicada.

Orientador: Prof. Dr. Maicon Rodrigues Albuquerque

**Belo Horizonte
2023**

2023, Ronaldo Angelo Dias da Silva.
Todos os direitos reservados

Silva, Ronaldo Angelo Dias da.

S586m Machine learning para a predição da carga de treinamento de um atleta da Seleção Brasileira de Taekwondo. Fenilcetonúria [recurso eletrônico] / Ronaldo Angelo Dias da Silva – 2023.

1 recurso online (58 f. il., color.) : pdf.

Orientador: Maicon Rodrigues Albuquerque

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 54-58

1. Estatística. 2. Controle preditivo 3. Tae kwon do – Treinamento técnico. 4. Inteligência artificial I. Albuquerque, Maicon Rodrigues. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz CRB
6/819 - Universidade Federal de Minas Gerais - ICEX




Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924


ATA DO 320ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE RONALDO ANGELO DIAS DA SILVA.

Aos quatorze dias do mês de dezembro de 2023, às 09:30 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Ronaldo Angelo Dias da Silva**, intitulado: “*Machine learning* para a predição da carga de treinamento de um atleta da seleção brasileira de *Taekwondo*.”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Guilherme Lopes de Oliveira Maicon Rodrigues de Albuquerque – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 14 de dezembro de 2023.

MAICON RODRIGUES
ALBUQUERQUE:0552185
4690
Assinado de forma digital por
MAICON RODRIGUES
ALBUQUERQUE:05521854690
Dados: 2023.12.18 15:46:09 -03'00'
Prof. Dr. Maicon Rodrigues de Albuquerque (Orientador)
EEFFTO/UFMG

Documento assinado digitalmente
 JOAO GUILHERME MAIA DE MENEZES
Data: 14/12/2023 14:20:02-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. João Guilherme Maia de Menezes
DCC/UFMG

Documento assinado digitalmente
 GUILHERME LOPES DE OLIVEIRA
Data: 14/12/2023 14:06:47-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Guilherme Lopes de Oliveira
DECOM/CEFET-MG

AGRADECIMENTOS

Ao meu orientador, Professor Dr. Maicon Rodrigues Albuquerque, por me guiar por esse caminho de desenvolvimento pessoal e profissional e, pela oportunidade única de aprendizado.

Aos meus professores da pós-graduação em Estatística Aplicada do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, por fornecerem a base necessária para eu pudesse percorrer esse árduo caminho.

Ao coordenador do curso de pós-graduação em Estatística Aplicada, o Professor Dr. Roberto da Costa Quinino, por quem eu tenho a mais elevada estima devido a todo o conhecimento transmitido. Muito obrigado por me incentivar para que eu pudesse prosseguir decididamente rumo ao meu desenvolvimento acadêmico.

EPÍGRAFE

”A imaginação é mais importante do que o conhecimento...”

(ALBERT EINSTEIN).

RESUMO

O objetivo do presente estudo foi ajustar um modelo preditivo para a carga de treinamento de um atleta da seleção brasileira de *Taekwondo*. Foram analisadas 122 sessões de treinamento e a escala de Bem_Estar, média de altura de salto com contramovimento (SCMmedia) e frequência cardíaca média (FC_MED) foram utilizadas como variáveis independentes. Para prever a percepção subjetiva do esforço em três níveis, Leve, Moderada e Elevada (PSE_class) foram ajustados os modelos de Árvore de Decisão de classificação, Floresta Aleatória, *Bagging* e *Boosting*. Após realizados os testes, a Árvore de Decisão apresentou acurácia de 26,09%, a Floresta Aleatória apresentou acurácia de 17,39%, o método *Bagging* apresentou acurácia de 56,52% e o *Boosting* demonstrou acurácia de 65,21%. Frente aos resultados encontrados, o método *Boosting* apresentou melhor desempenho e mostrou-se mais adequado para ajuste do modelo de árvore de decisão para prever a PSE_class de treinamento, mas deve ser usado com cautela.

Palavras Chave: Algoritmo de Predição; Inteligência Artificial; Análise Esportiva.

ABSTRACT

The aim of the present study was to adjust a predictive model for the training load of an athlete from the Brazilian Taekwondo team. 122 training sessions were analyzed and the Wellness_Being scale, countermovement jump height mean (CMJmean), and mean heart rate (HR_MED) were used as dependent variables. To predict the rate perception of exertion at three levels, Easy, Moderate and High (RPE_class), the classification Decision Tree, Random Forest, Bagging and Boosting models were adjusted. After carrying out the tests, the Decision Tree showed an accuracy of 26.09%, the Random Forest showed an accuracy of 17.39%, the Bagging method showed an accuracy of 56.52% and Boosting showed an accuracy of 65.21%. In conclusion, the Boosting method performed better and was more suitable for adjusting the decision tree model to predict the RPE_class of the training, but should be used with caution.

Keywords: Prediction Algorithm; Artificial Intelligence; Sports Analytics.

Lista de abreviaturas e siglas

AUC – área sob a curva ROC

Bagging – Bootstrap Aggregating

CART – Classification and Regression Trees

FC – Frequência Cardíaca

FC max – Frequência Cardíaca Máxima

FC_MED – Frequência Cardíaca Média

PSE – Percepção Subjetiva de Esforço

PSE class- Percepção Subjetiva de Esforço classificada em Leve, Moderada e Elevada

PSE Sessão - Percepção Subjetiva de Esforço da sessão de treinamento

PSE Treinador - Percepção Subjetiva de Esforço prevista pelo Treinador

PSR – Percepção Subjetiva de Recuperação

ROC – *Receiver Operating Characteristic*

SCM – Salto com contramovimento

SCMmedia – Média do salto com contramovimento

SCMsoma – Somatório do salto com contramovimento

Tempo Z1 – Tempo de treinamento com frequência cardíaca entre 65-71%

Tempo Z2 – Tempo de treinamento com frequência cardíaca entre 72-78%

Tempo Z3 – Tempo de treinamento com frequência cardíaca entre 78-85%

Tempo Z4 – Tempo de treinamento com frequência cardíaca entre 85-92%

Tempo Z5 – Tempo de treinamento com frequência cardíaca entre 92-100%

TRIMP – Impulso de Treinamento

Figuras

Figura 1. Particionamento de árvore de decisão	18
Figura 2. Matriz de confusão	23
Figura 3. Exemplo de Curva ROC.....	25
Figura 4. Percepção Subjetiva de Recuperação (PSR)	26
Figura 5. Fases do Salto com Contramovimento (SCM)	27
Figura 6. Percepção Subjetiva do Esforço da sessão (CR-10)	29
Figura 7. Árvore de decisão de classificação	37

Quadros

Quadro 1. Zonas de intensidade baseadas no percentual da frequência cardíaca.....	30
Quadro 2. Lista de variáveis	31
Quadro 3. Interpretação da árvore de decisão para classificação da PSE.	37
Quadro 4. Matriz de confusão da árvore de decisão.....	38
Quadro 5. Matriz de confusão da Floresta Aleatória para os dados de treinamento	41
Quadro 6. Matriz de confusão da Floresta Aleatória para os dados de teste.....	42
Quadro 7. Matriz de confusão do Bagging para os dados de teste.....	43

Gráficos

Gráfico 1. Matriz de Correlação	32
Gráfico 2. Distribuição de dados do Bem_Estar	35
Gráfico 3. Distribuição dos dados da FC_MED.....	35
Gráfico 4. Distribuição dos dados do SCMmedia	36
Gráfico 5. Distribuição e correlação das variáveis preditoras em diferentes classes.	36
Gráfico 6. Curva ROC da árvore de decisão para as diferentes classes	39
Gráfico 7. Seleção de hiperparâmetros para Floresta Aleatória	40
Gráfico 8. Taxa de erro out of bag da Floresta Aleatória.	40
Gráfico 9. Medida de importância das variáveis preditoras no modelo Floresta Aleatória	41
Gráfico 10. Curva ROC para modelo de Floresta Aleatória.....	42
Gráfico 11. Medida de importância das variáveis preditoras no modelo Bagging.....	44
Gráfico 12. Medida de importância das variáveis preditoras no modelo Boosting.....	44

Sumário

1 INTRODUÇÃO	14
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 <i>Machine Learning</i>	16
2.2 Árvore de decisão para regressão e classificação	16
2.3 <i>Overfitting</i>	18
2.4 <i>Bootstrap Aggregating (Bagging)</i>	19
2.5 <i>Boosting</i>	20
2.6 Floresta Aleatória (<i>Random Forest</i>)	21
2.7 Métricas de avaliação de desempenho do modelo	21
2.7.1 Avaliação de desempenho de modelos de regressão	21
2.7.2 Avaliação de desempenho de modelos de classificação	23
3 MATERIAIS E MÉTODOS	26
3.1 Participante e variáveis analisadas	26
3.2 Percepção Subjetiva de Recuperação (PSR)	26
3.3 Escala de Bem Estar (Escala HOOPER)	27
3.4 Salto com contramovimento (SCM)	27
3.5 Percepção Subjetiva do Esforço (PSE) da sessão (PSEsessão) e PSE classificada (PSEclass)	28
3.6 Frequência Cardíaca e Impulso de Treinamento (TRIMP)	29
3.7 Pré processamento dos dados e seleção de variáveis	30
3.8 Software utilizado	33
4 RESULTADOS	34
4.1 Análise exploratória das variáveis do modelo	34
4.2 Árvore de decisão CART	36
4.3 Floresta Aleatória	39
4.3 <i>Bootstrap Aggregating (Bagging)</i>	43
4.4 <i>Boosting</i>	44
4.5 Avaliação dos modelos	45
5 DISCUSSÃO	46
6 CONCLUSÃO	53
REFERÊNCIAS	54

1 INTRODUÇÃO

O treinamento esportivo é um processo complexo que visa aprimorar o desempenho, melhorar a capacidade física e reduzir o risco de lesões em atletas (Kiely, 2018). A prescrição de treinamento, antes baseada principalmente na experiência prática dos treinadores (Borresen e Lambert, 2009), atualmente se fundamenta em conceitos e métodos de diversas áreas, como fisiologia do exercício, biomecânica e psicologia do esporte (Vanrenterghem *et al.*, 2017). Assim, é possível perceber a complexidade de se prescrever treinamento e a natureza multifatorial do desempenho esportivo (Mujika *et al.*, 2018; Montull *et al.*, 2022).

Para elevar o nível de desempenho de um atleta é necessário submetê-lo a cargas ideais de treinamento (Furrer, Hawley e Handshin *et al.*, 2023). Portanto, as cargas não podem ser débeis, pois não proporcionarão melhora no desempenho e/ou excessivas, pois podem acarretar lesões (Gabbett, 2016). Por isso, a importância do monitoramento da carga de treinamento (Szmuchrowski e Couto, 2013).

O monitoramento da carga de treinamento é o processo de acompanhamento do treinamento por meio de ferramentas e tecnologias que quantificam o estímulo de treinamento e mensuram as respostas de fadiga/recuperação (McGuigan *et al.*, 2021). Sendo assim, o monitoramento da carga de treinamento tem como finalidades verificar o estado de prontidão do atleta para receber um novo estímulo, predizer o nível de desempenho, promover melhora no desempenho, diminuir o risco de lesões, evitar o *overreaching* não funcional e o *overtraining* (Bourdon *et al.*, 2017).

Diversas ferramentas e tecnologias têm sido empregadas para monitorar a carga de treinamento (Campbell *et al.*, 2017). Entre as mais utilizadas estão, a altura do salto com contramovimento - SCM (Claudino *et al.*, 2017), a percepção subjetiva do esforço (PSE) e percepção subjetiva do esforço da sessão - PSEsessão (Haddad *et al.*, 2017), a percepção subjetiva de recuperação - PSR (Laurent *et al.*, 2011), a escala de Bem-Estar - Hooper (Hooper e Mackinnon, 1995), a frequência cardíaca (Borresen e Lambert, 2008), o impulso de treinamento (TRIMP) (Fox *et al.*, 2018) dentre outros (Halsen, 2014). Entretanto, a interpretação dessas variáveis pode ser desafiadora (Foster, Marroyo e Koning, 2017; Schelling e Robertson, 2020), pois a relação complexa entre as diferentes métricas dificulta sua interpretação prática (Montull *et al.*, 2022; Helwig *et al.*, 2023). Portanto, é fundamental considerar a interdependência entre as variáveis e explorar métodos que permitam uma análise mais abrangente e precisa das informações

coletadas (Armstrong *et al.*, 2022). Nesse contexto, a utilização de técnicas de *Machine Learning* pode ser uma abordagem promissora para resolver esse problema (Didorf *et al.*, 2023).

São diversos os métodos empregados em *Machine Learning* para a construção de modelos estatísticos preditivos como, por exemplo, árvores de decisão de regressão e classificação, regressão linear e logística, *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM), *Random Forest* e *Gradient Boosting* (Hastie, Tibshirani e Friedman, 2009; Kassambara, 2017). Dentre esses modelos, árvores de decisão de classificação é o que possui mais fácil interpretabilidade (Hastie, Tibshirani e Friedman, 2009) e, por isso, podem auxiliar treinadores e comissão técnica no processo de tomada de decisão de ajuste da carga de treinamento a partir dos dados coletados no monitoramento.

A capacidade de identificar padrões complexos e realizar análises em grande escala torna o *Machine Learning* uma ferramenta importante para lidar com conjuntos de dados extensos e multifacetados (Chaudhary *et al.*, 2020; Rico-González *et al.*, 2023). Ao aplicar algoritmos de *Machine Learning*, podemos identificar relações entre as métricas de carga de treinamento e lesões no futebol (Rossi *et al.*, 2018), no basquetebol (Taborri *et al.*, 2021) e, em corredores (Rahlf *et al.*, 2022), realizar a predição de resultados esportivos no *Taekwondo* (Albuquerque *et al.*, 2021), prever o desempenho de atletas de voleibol (Leeuw *et al.*, 2022), detectar o tipo e velocidade do arremesso no handebol (van den Tillar, Bhandruga e Stewart, 2021) e determinar índices de desempenho físico em esportes de alto rendimento (Biró *et al.*, 2023). Portanto, essa abordagem pode permitir uma análise mais holística, considerando a interação entre as diversas variáveis envolvidas (Didorf *et al.*, 2023). Entretanto, trabalhos que utilizaram de técnicas de *Machine Learning* para auxiliar no monitoramento da carga de treinamento são escassos, principalmente nas modalidades esportivas de combate como o *Taekwondo*.

Embora a maior parte dos estudos utilizem o *Machine Learning* no esporte de alto rendimento com objetivo de prever lesões (Rommers *et al.*, 2020; Taborri *et al.*, 2021; Rahlf *et al.*, 2022; Ota e Kimura, 2023), o monitoramento da carga de treinamento pode oferecer informações importantes para as diferentes tomadas de decisões relacionadas ao planejamento e controle dos treinamentos (West *et al.*, 2020). Sendo assim, o objetivo do presente estudo é ajustar um modelo preditivo da carga de treinamento de um atleta da seleção brasileira de *Taekwondo* baseado em árvores de decisão de classificação.

2 FUNDAMENTAÇÃO TEÓRICA

As sessões 2.1, 2.2, 2.3, 2.4, 2.5, 2.6 e 2.7 a seguir são baseadas nas seguintes referências bibliográficas: (Nilsson, 1996; Mista, 2010; James *et al.*, 2013; Kuhn e Johnson, 2013; Hastie, Tibshirani e Friedman, 2009, Kassambara, 2017, Chaudhary *et al.*, 2020; Ferreira e Oliveira, 2020; Fávero, Belfiore e Souza, 2023).

2.1 *Machine Learning*

Machine Learning é uma subárea da inteligência artificial que utiliza algoritmos e modelos estatísticos para aprender padrões e tomar decisões a partir de dados. É amplamente aplicado em diversas áreas, permitindo a construção de modelos preditivos, classificatórios e de clusterização. São diversos os métodos empregados em *Machine Learning* para a construção de modelos estatísticos preditivos como, por exemplo, árvores de decisão, regressão linear e logística, *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM), *Random Forest* e *Gradient Boosting*.

2.2 **Árvore de decisão para regressão e classificação**

O método de árvore de decisão para regressão é uma técnica estatística amplamente utilizada em problemas de modelagem preditiva que envolvem variáveis contínuas como resposta. Já o método de as árvores de decisão de classificação, como o próprio nome já diz, é uma técnica de *machine learning* utilizada para resolver problemas de classificação, nos quais é necessário atribuir uma classe ou categoria a um determinado conjunto de dados com base em suas características. Ambos os métodos são conceituados como uma estrutura hierárquica de divisão de dados (*split*) baseada em condições lógicas aplicadas a variáveis independentes. A construção da árvore é realizada de forma iterativa e objetiva otimizar a divisão dos dados com base em algum erro de ajuste, que busca minimizar o desvio entre os valores previstos e os valores reais da variável resposta.

Um conceito fundamental em modelos de árvore de decisão, especialmente na seleção de critérios para dividir os nós da árvore, é o *Standard Deviation Reduction* (SDR). Ele representa a medida de quanto a variabilidade dos dados é reduzida após a divisão do nó, ou seja, quanto mais homogêneos os subgrupos resultantes da divisão, maior será a redução do desvio-padrão.

O método de árvore de decisão é particularmente atrativo devido à sua capacidade de lidar com não-linearidades e interações entre as variáveis independentes. Além disso, sua interpretabilidade é favorecida pela visualização da estrutura de ramificação da árvore,

permitindo uma compreensão clara dos fatores que influenciam as previsões. A inferência preditiva é geralmente conduzida por meio do cálculo da média ou moda das observações pertencentes ao conjunto de treinamento, que são associadas à região em que a nova observação foi classificada.

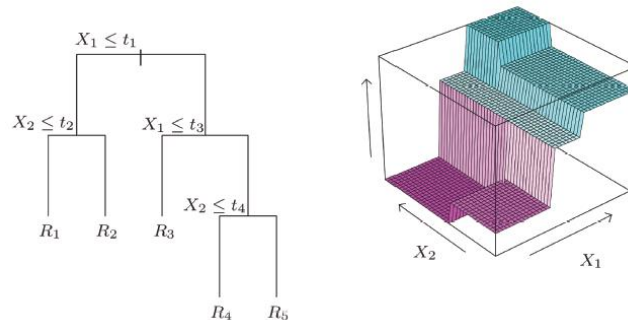
Em árvores de decisão, cada ramificação é chamada de nó. Sempre teremos apenas um nó raiz; podemos ou não ter nós de decisão; e teremos nós folha. As ramificações que conectam os nós entre si, do nó raiz a algum nó folha, são chamadas de ramos. Assim como as árvores na natureza, que iniciam seu crescimento a partir de suas raízes e desenvolvem um caule bem estruturado, galhos e depois folhas, as árvores de decisão também seguem essa ideia da biologia para serem construídas e compreendidas.

A construção da árvore de decisão para classificação e regressão é baseada em algoritmos que buscam encontrar o melhor particionamento dos dados em cada nó da árvore, considerando métricas de avaliação de qualidade, como o erro médio quadrático ou o coeficiente de determinação. Sendo assim, para a construção de árvores de regressão um dos métodos mais utilizados é o *Classification and Regression Trees (CART)* (Breiman *et al.*, 1984).

O método CART usa o conjunto de dados S , buscando os valores distintos de cada variável independente para encontrar, simultaneamente, a variável e o ponto de particionamento que minimizam a soma dos quadrados dos erros (SSE) nos grupos S_1 e S_2 . O procedimento é repetido para cada grupo, sendo conhecido como particionamento recursivo. A soma dos quadrados dos erros é obtida através da soma dos quadrados dentro de cada partição, em que y_i é o valor real de treinamento, \bar{y}_1 é o valor médio da primeira partição e \bar{y}_2 é o valor médio da segunda partição, de acordo com a equação a seguir: $ssE = \sum_{i \in S_1} (y_i - \bar{y}_1)^2 + \sum_{i \in S_2} (y_i - \bar{y}_2)^2$.

O particionamento obtido pela construção da árvore pode ser compreendido pela Figura 1. A árvore é construída em um espaço bidimensional das variáveis X_1 e X_2 . Os pontos de decisão t_1, \dots, t_4 guiarão novas observações para grupos R_1, \dots, R_5 , resultando em uma previsão média do conjunto de treinamento para cada grupo.

Figura 1. Particionamento de árvore de decisão



Fonte: Hastie, Tibshirani e Friedman (2009)

2.3 Overfitting

O *overfitting* em modelos estatísticos de árvores de decisão de regressão é caracterizado pela sobreajustagem do modelo aos dados de treinamento, ocasionando uma adaptação excessivamente precisa aos ruídos e variações aleatórias presentes no conjunto específico de treinamento. Como consequência, o modelo se torna altamente especializado para os dados de treinamento, mas sua capacidade de generalização para novos dados é prejudicada. Nesse contexto, o modelo pode apresentar um desempenho excepcional ao ajustar-se aos dados de treinamento, mas pode falhar ao lidar com dados não vistos durante o treinamento, gerando previsões menos precisas e confiáveis em situações reais.

Para minimizar o *overfitting* em modelos de árvores de decisão de regressão, são adotadas diversas estratégias que visam aprimorar a capacidade de generalização e a robustez das previsões. Estas abordagens incluem o controle de hiperparâmetros por meio da limitação da profundidade da árvore (*max_depth*) para reduzir sua complexidade, o estabelecimento de um número mínimo de amostras por folha (*min_samples_leaf*) para evitar subdivisões excessivamente especializadas, e a escolha minuciosa de critérios de divisão mais conservadores, como o erro quadrático médio (MSE), em detrimento de critérios que possam resultar em particionamentos excessivamente complexos.

A validação cruzada é outro processo que possibilita avaliar o desempenho do modelo em dados não observados, oferecendo a oportunidade de ajustar os hiperparâmetros de forma a minimizar o *overfitting*. Sendo assim, a incorporação de técnicas mais avançadas, como os métodos *Bagging* e *Random Forest* promovem uma maior estabilidade e generalização do modelo diante de variações nos dados de treinamento.

2.4 Bootstrap Aggregating (Bagging)

Árvores de decisão podem gerar modelos instáveis com previsões sensíveis a mudanças nos dados de treinamento. Para melhorar a estabilidade do modelo, pode-se utilizar o procedimento *Bootstrap Aggregating (bagging)*.

O processo de *Bagging* envolve a geração de várias réplicas do conjunto de treinamento original por meio de amostragem com reposição (*bootstrap*). Cada réplica é usada para treinar uma árvore de decisão independente. A amostragem com reposição permite que algumas observações sejam repetidas em cada subconjunto, enquanto outras podem ser excluídas. Na construção de preditores através do método *bagging*, em média dois terços das amostras do conjunto de dados são utilizados no treinamento de cada preditor. O restante das amostras recebe o nome *out-of-bag* (OOB).

Uma vez que as árvores de decisão são construídas em subconjuntos diferentes do conjunto de treinamento, elas podem apresentar variação nos resultados. No entanto, o *Bagging* agrega as previsões de todas as árvores através de uma média para chegar a uma única previsão final mais robusta e precisa.

Considerando um conjunto de n observações independentes, Z_1, \dots, Z_n , tendo uma variância σ^2 , a variância da média \bar{Z} das observações é dada por σ^2/n . Portanto, para reduzir a variância e aumentar a acurácia é necessário obter muitos conjuntos de treinamento, por meio de amostragem com reposição do conjunto de treinamento original. Assim, é possível construir vários modelos e obter a média das previsões $(\hat{f}^1(x), \hat{f}^2(x), \dots, \hat{f}^B(x))$, de B conjunto de treinamentos de acordo com a equação a seguir: $\hat{f}^{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x)$.

Essa abordagem reduz o risco de *overfitting*, pois as árvores individuais podem ser propensas a ajustar-se excessivamente aos dados de treinamento. Além disso, a técnica também pode capturar melhor as relações complexas entre as variáveis independentes e a variável de resposta, melhorando a capacidade de generalização do modelo e, conseqüentemente, sua acurácia em dados não vistos.

Uma das principais desvantagens do método *bagging* é que ele pode resultar em um aumento na variância das previsões. Embora o *bagging* seja eficaz em reduzir o *overfitting* e melhorar a precisão preditiva, a combinação de várias árvores de decisão independentes pode levar a previsões mais variáveis. Isso ocorre porque as árvores individuais podem ser altamente

correlacionadas, especialmente quando há um preditor muito forte no conjunto de dados. Como resultado, as previsões das árvores agrupadas podem estar mais próximas entre si do que das previsões verdadeiras, resultando em uma estimativa da variabilidade que pode não refletir adequadamente a incerteza dos dados. Para minimizar essa desvantagem, são empregados métodos como o *Random Forest*, que introduz uma técnica adicional de aleatorização para descorrelacionar as árvores e reduzir a variância das previsões.

2.5 Boosting

O Boosting é uma técnica de *Machine Learning* introduzida nos últimos vinte anos que visa melhorar a performance dos modelos de árvores de decisão, expandindo sua capacidade de generalização por meio da combinação de múltiplas árvores. Essa técnica pode ser dividida em diversas etapas, em que inicialmente, uma árvore de decisão fraca é treinada no conjunto de dados original. Os exemplos de treinamento são ponderados, de modo que os exemplos classificados erroneamente tenham maior influência. Sendo assim, com base nos erros cometidos pela primeira árvore, uma segunda árvore é construída para corrigir esses erros. Esse processo é repetido sequencialmente, onde cada nova árvore tem o objetivo de melhorar a precisão e generalização das anteriores.

As equações fundamentais no *Boosting* incluem a atualização dos pesos dos exemplos e a combinação das previsões das árvores individuais. A atualização dos pesos é realizada para dar maior importância aos exemplos classificados incorretamente, incentivando as árvores subsequentes a focarem nesses casos. A equação de atualização dos pesos é dada por, $\omega_i^{(m+1)} = \omega_i^{(m)} \cdot e^{-\alpha_m y_i h_m(x_i)}$, onde $\omega_i^{(m)}$ é o peso do exemplo i na interação m , α_m é o coeficiente de aprendizado na interação m , y_i é o rótulo do exemplo i e $h_m(x_i)$ é a previsão da árvore m para o exemplo i .

A combinação das previsões é realizada ponderando as previsões individuais das árvores de acordo com seus coeficientes de aprendizado. A previsão final é dada por: $FM(x) = \sum_{m=1}^M \alpha_m h_m(x)$, onde $FM(x)$ é a previsão final do modelo *Boosting*, M é o número de árvores e $h_m(x)$ é a previsão da árvore m para o exemplo x . Sendo assim, as equações fundamentais do *Boosting* refletem o processo de atualização dos pesos e a combinação das previsões, demonstrando a natureza iterativa e colaborativa do método.

2.6 Floresta Aleatória (*Random Forest*)

As Florestas Aleatórias representam um tipo de modelos baseado em *ensemble methods*, os quais consistem na combinação de predições de múltiplas árvores de decisão, resultando em uma resposta única como. Essa abordagem é conhecida por proporcionar um melhor desempenho em relação ao uso individual de cada árvore do modelo, devido à redução de variância. Através da agregação das previsões de várias árvores independentes, as Florestas Aleatórias têm a capacidade de capturar relações complexas entre as variáveis e produzir resultados mais robustos e precisos em tarefas de *Machine Learning*. Essa técnica é amplamente empregada em diversos problemas práticos, tornando-se uma estratégia eficaz na construção de modelos preditivos mais confiáveis.

As florestas aleatórias fornecem uma melhoria em relação às árvores agregadas por meio de um pequeno ajuste chamado "*Random Forest*" que descorrelaciona as árvores. Esse ajuste consiste em criar múltiplas árvores independentes e descorrelacionadas, reduzindo a dependência entre elas e aumentando a diversidade das previsões.

2.7 Métricas de avaliação de desempenho do modelo

2.7.1 Avaliação de desempenho de modelos de regressão

Para avaliar os modelos de regressão podemos utilizar diversas métricas. A mais empregada é a raiz do erro quadrático médio, o *root mean square error* (RMSE), que mede a diferença média entre os valores observados e os valores previstos pelo modelo, sendo uma medida da dispersão dos erros de previsão.

Nos modelos de árvore de decisão, cada árvore individual do *Random Forest* faz suas previsões para o conjunto de dados de teste. O RMSE (equação 5) é calculado para cada árvore individual, encontrando a diferença entre os valores reais e as previsões de cada árvore, elevando essas diferenças ao quadrado e tirando a média desses quadrados. Por fim, o RMSE final do *Random Forest* é calculado tirando a raiz quadrada da média dos RMSEs das árvores individuais, dado

pela equação, $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$. O RMSE é uma métrica importante porque fornece uma visão geral da precisão das previsões do modelo. Quanto menor o valor do RMSE, melhor é o ajuste do modelo aos dados de teste.

Outra métrica utilizada para avaliação de modelos de árvore de decisão é o erro médio absoluto (*Mean Absolute Error* - MAE) que mensura discrepância média entre os valores previstos pelo

modelo e os valores reais, por meio da soma das diferenças absolutas entre esses dois conjuntos de valores e sua subsequente divisão pelo número de observações, de acordo com a equação,

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i).$$

Para árvores de decisão, o MAE é aplicado de modo que cada árvore individual efetua previsões para um conjunto de dados de teste e, subsequentemente, o MAE é calculado para cada árvore, refletindo a média das discrepâncias absolutas entre as previsões realizadas pela árvore e os valores reais correspondente. Aferindo-se o MAE final, por meio da média dos MAEs das árvores individuais, é possível avaliar o ajuste global do modelo aos dados de teste, com valores menores indicando uma concordância mais acurada entre as previsões e os valores observados.

A métrica de erro médio absoluto percentual, conhecida como *Mean Absolute Percentage Error* (MAPE) é utilizada para medir o erro médio percentual entre os valores reais e os valores previstos pelo modelo. O MAPE é usado para avaliar o quão bem as previsões do modelo estão em termos percentuais, o que pode ser útil para compreender o impacto das discrepâncias em diferentes escalas de valores. Em outras palavras, a interpretação do resultado do MAPE envolve analisar a média das diferenças percentuais absolutas entre as previsões e os valores reais. Um MAPE menor indica que as previsões estão mais próximas dos valores reais em termos percentuais. A equação 7 a seguir demonstra como é calculado o MAPE, dado pela

$$equação, MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100.$$

Por fim, o coeficiente de determinação, também conhecido como R^2 , é uma métrica estatística frequentemente utilizada para avaliar a qualidade e a capacidade de explicação de um modelo de regressão. Ele mede a proporção da variabilidade dos valores da variável dependente que é explicada pelas variáveis independentes do modelo (Ferreira e Oliveira, 2020). O valor do R^2 varia de 0 a 1, onde um valor mais próximo de 1 indica que o modelo é capaz de explicar uma porção maior da variabilidade dos dados, calculado pela equação, $R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$.

2.7.2 Avaliação de desempenho de modelos de classificação

Para avaliar a capacidade discriminatória de um modelo de árvore de decisão de classificação utilizam-se métricas como, por exemplo, a acurácia, precisão, sensibilidade, especificidade, F1 score e acurácia balanceada, que são calculadas a partir de uma matriz de confusão. Uma matriz de confusão é uma representação tabular que condensa os resultados das previsões feitas por um modelo em relação aos valores reais das classes em um conjunto de dados. Essa matriz permite uma análise sistemática da performance do modelo, discriminando as previsões corretas (verdadeiros positivos e verdadeiros negativos) das previsões incorretas (falsos positivos e falsos negativos). A figura 2 a seguir apresenta uma matriz de confusão.

Figura 2. Matriz de confusão

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: google imagens

A acurácia é uma métrica amplamente utilizada na avaliação de modelos de classificação e representa a proporção de previsões corretas feitas pelo modelo em relação ao número total de previsões e é fornecida pela equação, $Acurácia = \frac{TP+TN}{TP+FP+TN+FN}$. Apesar de ser uma métrica usual e de fácil interpretação, ela pode não ser a mais adequada e deve-se tomar cuidado ao levá-la em consideração em casos que os dados são desbalanceados, pois o modelo pode aprender a classificar melhor a classe que mais aparece do que a classe que está em menor proporção, dessa forma inflando o valor da acurácia.

A precisão é uma métrica que quantifica a capacidade do modelo em não classificar como positiva uma observação que na realidade é negativa, ou seja, de todas as previsões positivas, quantas estão corretas. A precisão é calculada a partir, $Precisão = \frac{TP}{TP+FP}$.

A sensibilidade, também conhecida como *recall*, avalia a capacidade do modelo em identificar corretamente todos os casos positivos presentes nos dados. Sendo assim, a sensibilidade mede a proporção de exemplos verdadeiros positivos (ou seja, casos em que o modelo previu corretamente a classe positiva) em relação ao total de casos positivos reais no conjunto de dados. Uma alta sensibilidade indica que o modelo tem uma capacidade sólida de identificar

positivos verdadeiros, minimizando os falsos negativos (casos positivos que foram incorretamente classificados como negativos). O cálculo da sensibilidade é dado pela equação,

$$\text{Sensibilidade} = \frac{TP}{TP+FN}.$$

A especificidade é uma medida que se concentra na capacidade do modelo em identificar corretamente todos os casos negativos presentes nos dados. Sendo assim, a especificidade mede a proporção de exemplos verdadeiros negativos (ou seja, casos em que o modelo previu corretamente a classe negativa) em relação ao total de casos negativos reais no conjunto de dados. Uma alta especificidade indica que o modelo tem a habilidade de evitar a classificação incorreta de casos negativos como positivos, minimizando os falsos positivos (casos negativos que foram incorretamente classificados como positivos). O cálculo da especificidade é dado pela equação,

$$\text{Especificidade} = \frac{TN}{TN+FP}.$$

O F1 score é uma medida que combina precisão e recall em uma única métrica. Ele busca encontrar um equilíbrio entre essas duas métricas, sendo particularmente útil quando as classes estão desbalanceadas no conjunto de dados. Um valor mais alto de F1 score indica um melhor equilíbrio entre a capacidade do modelo de fazer previsões precisas e sua habilidade de identificar todos os casos positivos. A equação que define o cálculo do é dado por, $F1 \text{ score} =$

$$\frac{2 \times \text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}.$$

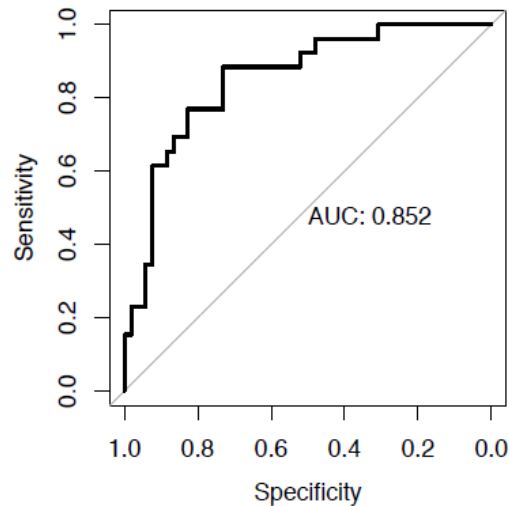
A acurácia balanceada é uma métrica de desempenho que leva em consideração o desbalanceamento das classes no conjunto de dados. Enquanto a acurácia tradicional mede a proporção geral de previsões corretas, a acurácia balanceada calcula a média das sensibilidades de cada classe presente no problema de classificação. Isso significa que a acurácia balanceada dá peso igual a todas as classes, tornando-a especialmente útil quando o desbalanceamento entre as classes é significativo. Ela fornece uma visão mais precisa do desempenho do modelo, evitando que a acurácia seja inflada por uma classe dominante. Portanto, a acurácia balanceada é uma métrica valiosa para avaliar modelos de árvore de decisão em situações em que a

distribuição das classes é desigual. O cálculo é dado por, $\text{Acurácia Balanceada} = \frac{1}{2} \frac{TP}{TP+FN} + \frac{1}{2} \frac{TN}{TN+FP}.$

Além das medidas reportadas anteriormente, a curva ROC (*Receiver Operating Characteristic*) é uma ferramenta gráfica de avaliação usada para medir o desempenho de modelos de classificação. Ela representa a taxa de verdadeiros positivos no eixo y (sensibilidade) e a taxa

de falsos positivos no eixo x (especificidade). A figura 3 a seguir apresenta o exemplo uma curva ROC.

Figura 3. Exemplo de Curva ROC



Legenda: Sensitivity: Sensibilidade; Specificity: Especificidade; AUC: área sob a curva ROC

Fonte: Kassambara, 2017

A curva ROC é construída variando o limite de decisão do modelo de classificação e calculando a taxa de sensibilidade e a taxa de especificidade para cada limite. Isso gera um conjunto de pontos na curva ROC. Um modelo perfeito teria uma curva ROC que passa pelo canto superior esquerdo, enquanto um modelo que faz previsões aleatórias teria uma curva ROC que é uma linha reta diagonal. Além disso, a área sob a curva (AUC) é frequentemente usada para resumir o desempenho geral do modelo de classificação. Quanto maior a AUC, melhor o modelo é em distinguir entre as classes. Uma AUC de 0,5 representa um modelo que não é melhor do que previsões aleatórias, enquanto uma AUC de 1 representa um modelo perfeito.

Por fim, o teste de Kappa avalia a concordância entre as previsões do modelo e as classes reais, corrigindo-a pela concordância esperada ao acaso. Isso fornece uma medida mais robusta do desempenho do modelo do que simplesmente observar a taxa de acerto. Sendo assim, é particularmente útil quando se lida com classes desbalanceadas. O valor do Kappa varia de -1 a 1, onde 1 indica uma concordância perfeita, 0 indica concordância igual ao acaso e valores negativos indicam uma concordância inferior ao acaso.

3 MATERIAIS E MÉTODOS

3.1 Participante e variáveis analisadas

Participou do presente estudo um atleta da seleção Brasileira de *Taekwondo* que possui experiência em competições nacionais e internacionais. Os dados coletados são referentes ao período de 03/01/2022 a 29/11/2022 e, foram analisadas 122 sessões de treinamento. Neste sentido, observações referentes à duração da sessão de treinamento, à PSR, à escala de Bem-Estar, o teste de SCM, à PSE, à PSEsessão, frequência cardíaca média e máxima e o TRIMP e suas subdivisões de tempo em cada uma das cinco zonas de intensidade de treinamento foram coletadas. Os dados utilizados neste projeto foram disponibilizados pelo Centro de Treinamento Esportivo da UFMG.

3.2 Percepção Subjetiva de Recuperação (PSR)

Com intuito de se monitorar o estado de recuperação de atletas foi proposto a Percepção Subjetiva de Recuperação (PSR) que, assim como a Percepção Subjetiva do Esforço é uma variável psicofísica que possui 11 graduações (0 a 10) (Laurent *et al.*, 2011).

A aplicação do PSR é bastante simples, para tal, basta que o atleta indique numa tabela qual é seu estado de recuperação em relação à sessão de treinamento anterior. O avaliador deve instruir o avaliado a escolher um descritor e depois um número de 0 a 10. O valor máximo (10) deve ser comparado a totalmente recuperado e o valor mínimo (0) é a condição em nada recuperado. A Figura 4 a seguir apresenta a PSR seus valores e descritores.

Figura 4. Percepção Subjetiva de Recuperação (PSR)

Selecione na escala a seguir o quão recuperado você se sente

- 0 - Muito mal recuperado - extremamente cansado
- 1
- 2 - Não bem recuperado - um pouco cansado
- 3
- 4 - Um pouco recuperado
- 5 - Adequadamente recuperado
- 6 - Moderadamente recuperado
- 7
- 8 - Bem recuperado - um pouco energético
- 9
- 10 - Muito bem recuperado - altamente energético

Fonte: adaptado de Laurent *et al.* (2011)

3.3 Escala de Bem Estar (Escala HOOPER)

O índice de Hooper avalia quatro constructos nos domínios de fadiga, estresse, dor muscular e sono. O objetivo da utilização do questionário de bem-estar é identificar possíveis sinais e sintomas relacionados à sobrecarga do treinamento que podem indicar estagnação, melhora ou queda de desempenho. Sendo assim, o atleta aponta em uma escala tipo likert que vai de 1 (muito, muito baixo, ou bom) a 7 (muito, muito alto, ou ruim) qual a sua percepção em relação a esses domínios (Hooper *et al.*, 1995). Posteriormente, o bem-estar geral foi determinado pela soma das pontuações (Hooper *et al.*, 1995). Quanto maior o valor do somatório, pior é o estado de Bem_Estar do atleta.

3.4 Salto com contramovimento (SCM)

O SCM tem sido utilizado nas mais diversas modalidades esportivas de combate como, o judô (Franchini *et al.*, 2011), o *Brazilian jiu-jitsu* (Andreato *et al.*, 2017), *Kickboxing* (Ouergui *et al.*, 2015) e, no próprio *Taekwondo* (Silva *et al.*, 2020), para estimar a altura do salto e monitorar a carga de treinamento (Claudino *et al.*, 2017).

O teste consiste na execução de um salto a partir da posição ortostática, com joelhos estendidos e as mãos apoiadas no quadril, na região supra-ilíaca (Gheller *et al.*, 2014). O voluntário executa uma ação excêntrica de flexão de joelhos até a angulação que julgar mais eficiente seguida por uma ação concêntrica de extensão de joelhos (Gheller *et al.*, 2014). Os joelhos devem permanecer estendidos durante a fase de voo e, a aterrissagem deverá ocorrer em flexão plantar (Gheller *et al.*, 2014). A Figura 5 a seguir apresenta o SCM.

Figura 5. Fases do Salto com Contramovimento (SCM)



Fonte: google imagens

O processo de familiarização seguiu os procedimentos sugeridos por Szmuchrowski *et al.* (2012). Anterior à realização dos testes de SCM, os voluntários realizaram o aquecimento descrito por Silva *et al.* (2023) no qual o indivíduo, por 5 minutos, realizava uma corrida de baixa intensidade (trote), fazia *sprints* curtos ($\approx 10\text{m}$).

Para análise de dados do SCM foram utilizadas a variável de altura de salto em centímetros ($SCM_{(cm)}$) (Gathercole *et al.*, 2015). Essa variável é discriminante para diferentes níveis de competidores, nacionais e internacionais (Bridge *et al.*, 2014; Santos *et al.*, 2020). Além disso, essa medida apresenta correlação forte e positiva com a velocidade de chute do *bandal tchagui* (Goulart *et al.*, 2016). Sendo assim, foram realizados três saltos (SCM1, SCM2 e SCM3) e calculado a média (SCMmedia) e o somatório (SCM soma) dos três saltos.

3.5 Percepção Subjetiva do Esforço (PSE) da sessão (PSEsessão) e PSE classificada (PSEclass)

A percepção subjetiva do esforço (PSE) é uma variável psicofísica que responde à integração de sinais aferentes de diversos sistemas durante o exercício, como da musculatura e articulações periféricas, do sistema nervoso central, cardiovascular e respiratório (Borg, 1982). As escalas mais conhecidas para sua mensuração são a escala PSE de com 15 graduações (6 a 20) e a escala de relações de categorias (CR10) e, sua utilização se baseia na classificação contínua do indivíduo com o aumento crescente e gradual da PSE do exercício (Kautzner, 2013). Além disso, alguns estudos tem proposto a utilização da CR10 em subdivisões para classificação da PSE (PSEclass) em Leve ($CR10 \leq 3$), Moderada ($CR10 \geq 4$ e ≤ 7) e Elevada ($CR10 > 7$) (Inoue *et al.*, 2022).

O método da PSE da sessão foi proposto com intuito de quantificar a carga de treinamento levando em consideração não só a intensidade, mas também a duração da sessão (Foster *et al.*, 2001). Sendo assim, trinta minutos após o término da sessão de treino, o atleta deve responder à seguinte pergunta: “Como foi a sua sessão de treino?” A resposta ao questionamento é fornecida a partir da escala. A PSE sessão é obtida multiplicando a duração de cada sessão de treinamento (em min) pela intensidade atribuída a essa sessão na escala (Foster *et al.*, 2001).

Este método tem se destacado na literatura por sua fácil aplicação e baixo custo operacional (Borresen & Lambert, 2008). O produto da multiplicação entre a duração da sessão e intensidade correspondente a percepção de esforço do atleta representa o valor da magnitude da carga expresso em unidades arbitrárias [UA] (Foster *et al.*, 2001). A figura 6 a seguir apresenta um modelo da PSE utilizada.

Figura 6. Percepção Subjetiva do Esforço da sessão (CR-10)

Nota	Descritor
0	Repouso
1	Muito, muito fácil
2	Fácil
3	Moderado
4	Um pouco difícil
5	Difícil
6	-
7	Muito difícil
8	-
9	-
10	Máximo

Fonte: Borg (1982), modificada por Foster *et al.*, (1996)

A PSE sessão foi calculada a partir da equação proposta por Foster *et al.* (1996), $PSE_{sessão} = PSE (CR-10) \times \text{duração de treino (minutos)}$

A PSE também pode ser utilizada pelo treinador, que estipula uma intenção de PSE a ser alcançada em determinada sessão e/ou período do treinamento (Paul *et al.*, 2021). Posteriormente a PSE esperada pelo treinador é comparada com a PSE reportada pelo atleta para verificação da adequação da carga de treinamento aplicada (Paul *et al.*, 2021).

Estudos demonstram que há concordância entre treinadores e atletas sobre a PSE e o PSE sessão, bem como a PSE e a PSE sessão divididos em duas categorias de esforço (moderado e intenso) (Inoue *et al.*, 2022). No entanto, houve discordância na PSE e na PSE sessão para a categoria de esforço Leve. Sendo assim, apesar de pequenos desacordos, o uso dessas ferramentas parece ser adequado para o monitoramento do treinamento (Inoue *et al.*, 2022).

3.6 Frequência Cardíaca e Impulso de Treinamento (TRIMP)

O Impulso de Treinamento, ou TRIMP (do inglês "*Training Impulse*"), é uma métrica utilizada para quantificar a carga interna de treinamento durante uma sessão de treinamento ou um período de tempo específico (Borresen e Lambert, 2008b). Essa métrica foi desenvolvida para avaliar de forma mais precisa a intensidade e a duração do exercício, levando em consideração não apenas o tempo gasto, mas também a frequência cardíaca e a zona de treinamento em que o atleta está (McGuigan *et al.*, 2021).

O TRIMP é calculado multiplicando a duração do exercício pelo valor da frequência cardíaca média elevado a um fator de intensidade (Borresen e Lambert, 2008b). Essa constante, fator de

intensidade, é obtida por meio de um teste de esforço, pode variar de acordo com a intensidade do treinamento e pode ser ajustada com base na experiência e nas características do atleta (McGuigan *et al.*, 2021).

Para esportes de característica intermitente, como o caso do Taekwondo (Silva *et al.*, 2023), a TRIMP modificada (TRIMPmod) é a mais recomendada por apresentar melhor ajuste da frequência cardíaca (Stagno, Thatcher e Someren, 2007). A TRIMPmod é utilizada para avaliar o volume e a intensidade da sessão de treinamento através de escores específicos determinada por zonas baseada no percentual da frequência cardíaca máxima que é registrada durante os períodos monitorados (Stagno, Thatcher e Someren, 2007). O quadro 1 a seguir apresenta as zonas de treinamento.

Quadro 1. Zonas de intensidade baseadas no percentual da frequência cardíaca

Zona	% frequência cardíaca máxima	Fator de intensidade
1	65% - 71%	1,25
2	72% - 78%	1,71
3	79% - 85%	2,54
4	86 % - 92%	3,61
5	93% - 100%	5,16

O tempo que o atleta permanece em cada uma das zonas é multiplicado pelo respectivo fator de ponderação e realizado o somatório (Stagno, Thatcher e Someren, 2007). Para determinar a frequência cardíaca e o tempo em cada zona de treinamento, utiliza-se um transmissor cardíaco (POLAR®, modelo H10 - Kempele, Finlândia) e os dados foram extraídos do *software* Polar Team.

3.7 Pré processamento dos dados e seleção de variáveis

Inicialmente, foi realizada uma análise descritiva dos dados referente aos valores mínimos e máximos, média, desvio-padrão, mediana, distância interquartil, dados faltantes e *out liers*. A princípio, nenhum valor discrepante foi removido do banco de dados por se tratar de dados coletados de um único atleta. Também não foram registrados dados faltantes. As variáveis qualitativas nominais foram automaticamente removidas da análise. Para as demais variáveis, o critério de exclusão levou em consideração valores elevados de correlação (gráfico 1) entre variáveis, praticidade na aplicação do teste para coleta de dados e, a interpretabilidade das medidas. Após essa análise, foi definida como variável a ser predita a PSEclass devido a sua

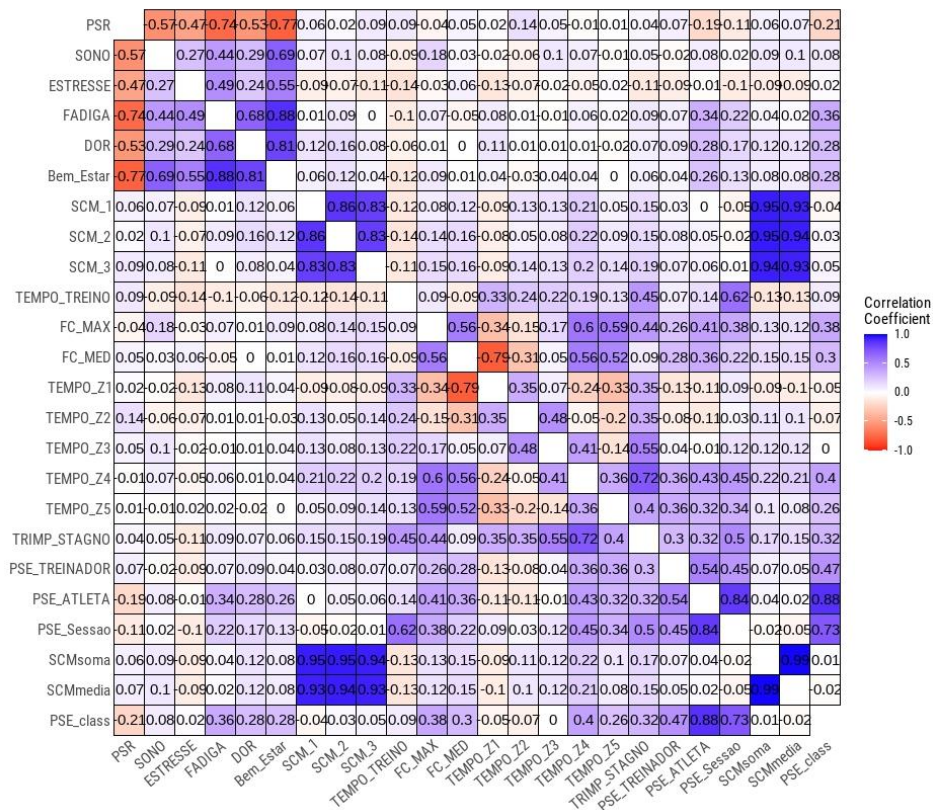
mais fácil aplicação e interpretabilidade. Sendo assim, as demais variáveis foram definidas como preditoras. O quadro 2 a seguir apresenta a lista das principais variáveis.

Quadro 2. Lista de variáveis

Tipo	Subtipo	Nº	Variável
Qualitativa	Nominal	01	ID
		02	Nome
		03	Data
		04	Dia
	Ordinal	05	PSR
		06	Sono
		07	Estresse
		08	Fadiga
		09	Dor
		10	Bem_Estar
		11	PSE Atleta
		12	PSE Sessão
		13	PSE Treinador
		14	PSEclass
Quantitativa	Contínua	15	SCM1
		16	SCM2
		17	SCM3
		18	SCMmedia
		19	SCM soma
		20	FCmax
		21	FCmed
		22	Tempo Treino
		23	Tempo Z1
		24	Tempo Z2
		25	Tempo Z3
		25	Tempo Z4
		27	Tempo Z5
		28	TRIMP

Legenda: ID: código identificador; Nome: nome completo do atleta; Data: dia mês e ano; Dia: dia da semana; PSR: percepção subjetiva de recuperação; PSE: percepção subjetiva do esforço; PSE sessão: percepção subjetiva do esforço multiplicada pela duração da sessão; PSE treinador: intenção de percepção subjetiva do esforço do treinador; PSE class: classificação da percepção subjetiva do esforço em Leve, moderada e Elevada; SCM1: primeiro salto com contramovimento; SCM2: segundo salto com contramovimento; SCM3: terceiro salto com contramovimento; SCMmedia: média dos SCM1, SCM2 e SCM3; SCM soma: somatório do SCM1, SCM2 e SCM3; FCmax: frequência cardíaca máxima; FCmed: frequência cardíaca média; Tempo Treino: duração da sessão de treinamento em minutos; Tempo Z1: tempo de treinamento com frequência cardíaca entre 65% - 71% do máximo; Tempo Z2: tempo de treinamento com frequência cardíaca entre 72% - 78% do máximo; Tempo Z3: tempo de treinamento com frequência cardíaca entre 79% - 85% do máximo; Tempo Z4: tempo de treinamento com frequência cardíaca entre 86% - 92% do máximo; Tempo Z5: tempo de treinamento com frequência cardíaca entre 93% - 100% do máximo; TRIMP: impulso de treinamento.

Gráfico 1. Matriz de Correlação



Analisando o gráfico 1, é possível perceber valores elevados de correlação entre as variáveis PSR, Sono, Estresse, Fadiga, Dor e Bem_Estar. Sendo assim, por entender que a o Bem_Estar deriva das informações de Sono, Estresse, Fadiga, Dor que influenciam diretamente na percepção de recuperação do atleta, o Bem_Estar foi mantido para análise, enquanto as demais variáveis foram excluídas.

O mesmo procedimento foi adotado para o SCM1, SCM2, SCM3, SCMmedia e SCM soma que apresentaram elevados valores de correlação (gráfico 1). Sendo assim, por entender que o SCMmedia contém as informações do SCM1, SCM2, SCM3 e, por isso representar o comportamento típico do atleta, além de ser de mais fácil interpretabilidade do que o SCM soma, este foi mantido para análise ao passo que os demais foram excluídos.

Os Tempos em Z1, Z2, Z3, Z4 e Z5 são de difícil quantificação em tempo real e possuem valores elevados de correlação com a TRIMP, que por sua vez possui elevados valores de correlação com a FC média e FC máxima. Sendo assim, devido à praticidade de coleta e interpretabilidade do dado, optou-se por manter a FC média e excluir as demais variáveis juntamente com a FC máxima que também possui elevada correlação com a FC média.

A PSE reportada pelo atleta utilizando a escala CR10 possui valor elevado de correlação com a PSE sessão que por sua vez possui valor elevado de correlação com o Tempo Treino. Portanto, tanto PSE e PSE sessão foram excluídas da análise. A PSE do treinador foi retirada do modelo, pois trata-se do único que não é decorrente de teste realizado pelo atleta. Sendo assim, a PSE_class foi mantida por se tratar de uma métrica que representa todas as outras e por possuir mais fácil interpretabilidade para análise da carga de treinamento em três níveis, “Leve”, “Moderada e “Elevada”.

Em suma, das 24 variáveis possíveis para o desenvolvimento do modelo, foram mantidas 3 predictoras ou independentes (o Bem_Estar, o SCMmedia e a FC_MED). A PSEclass foi mantida como variável a ser predita ou dependente. A escolha por essas variáveis teve como base critérios estatísticos e teóricos que levam em consideração as inúmeras possibilidades de monitoramento da carga de treinamento, por meio de parâmetros mecânicos, fisiológicos e psicológicos (Vanrenterghem *et al.*, 2017; Mujika *et al.*, 2018; Montull *et al.*, 2022).

3.8 Software utilizado

Para o desenvolvimento deste trabalho foi empregada a linguagem R usando o *software* Rstudio versão 2023.06.1-524. Os pacotes utilizados foram: O *dlookr* (Ryu Choonghyun, 2023) para o pré processamento dos dados. O *Modelsummary* (Arel-Bundock, 2022) para relatório diagnóstico. O *Caret* (Kuhn, 2008) para criação de conjunto de treinamento e teste do modelo, permitindo comparar as combinações de hiperparâmetros utilizando validação cruzada. O *rpart* (Therneau, Atkinson e Ripley, 2017) e *rpart.plot* (Milborrow, 2019) para construção e visualização de árvores de decisão. O *gbm* (Greenwell, Boehmke e Cunningham, 2020) para melhorar a estabilidade do modelo, por meio de *Bootstrap Aggregating (bagging)* e o *xgboost* (Chen e Guestrin., 2016) para realização do *Boosting*.

4 RESULTADOS

4.1 Análise exploratória das variáveis do modelo

Do conjunto total de 122 observações, foram particionados aleatoriamente 99 (80%) para treinamento e 23 (20%) para teste. Na Tabela 1 a seguir são apresentadas algumas estatísticas descritivas a respeito das variáveis do conjunto total de dados. O gráfico 2 a seguir apresenta a distribuição do Bem_Estar. O gráfico 3 apresenta a distribuição dos dados da FC_MED, o gráfico 4 a seguir apresenta a distribuição dos dados do SCMmedia e, o gráfico 5 a seguir apresenta os valores de correlação das variáveis preditoras e suas respectivas distribuições em relação à variável predita em suas diferentes classes.

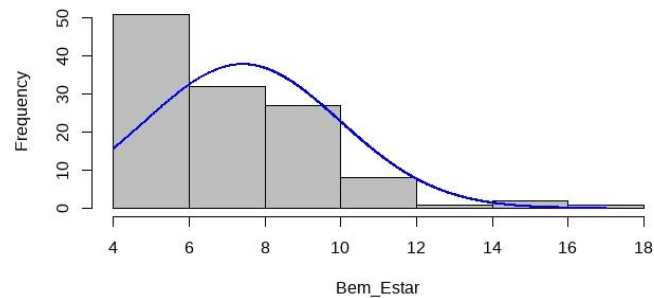
Tabela 1. Análise descritiva do conjunto total de dados

Variável	Mínimo e Máximo	Q1 Mediana Q2	Média e DP	EPM	IQR	Assimetria	Curtose
Bem_Estar (u.a)	4 – 17	6 – 7 – 9	7,41 ± 2,58	0,24	3	0,96	1,54
FC_MED (bpm)	84 – 161	119 – 129 – 138	128,00 ± 14,40	1,31	19	- 0,28	0,17
SCMmedia (cm)	25 – 43	37 – 38 – 39	38,30 ± 2,18	0,20	2	- 1,64	10,40

Legenda: Q1: primeiro quartil; Q2: segundo quartil; DP: desvio padrão; EPM: erro padrão da média; IQR: intervalo interquartil; PSR: Percepção subjetiva de recuperação; FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento.

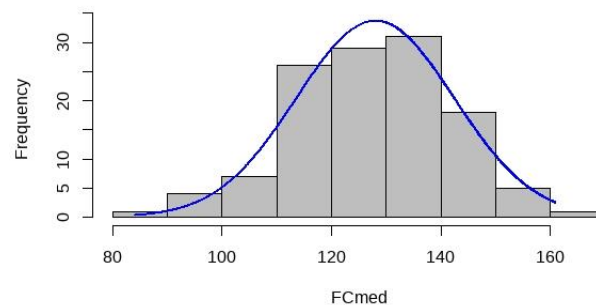
O Bem_Estar demonstra distribuição de dados levemente assimétrica à direita. Esse resultado sugere que a cauda direita da distribuição é um pouco mais longa que a cauda esquerda, implicando que valores mais altos de bem-estar são menos frequentes, resultando em uma distribuição não simétrica. Além disso, a curtose positiva denota uma distribuição leptocúrtica, caracterizada por caudas mais pesadas e um pico mais acentuado em comparação com a distribuição normal. O gráfico 2 a seguir apresenta a distribuição do Bem_Estar.

Gráfico 2. Distribuição de dados do Bem_Estar



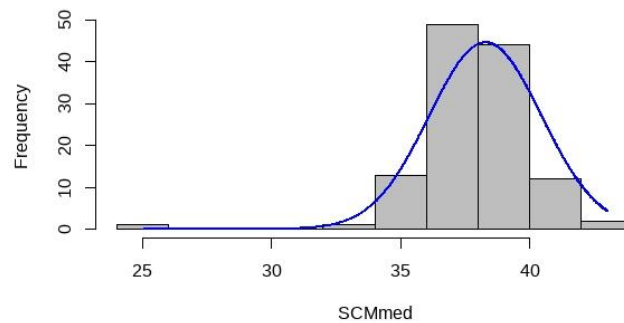
Para a variável FC_MED apresenta uma leve assimetria à esquerda na distribuição. Isso sugere que valores de frequência cardíaca média mais baixos são menos frequentes, resultando em uma cauda esquerda ligeiramente mais longa. Além disso, a curtose próxima de zero sugere que a distribuição é quase normal, com um formato geral próximo ao de uma distribuição gaussiana. O gráfico 3 a seguir apresenta a distribuição dos dados da FC_MED.

Gráfico 3. Distribuição dos dados da FC_MED



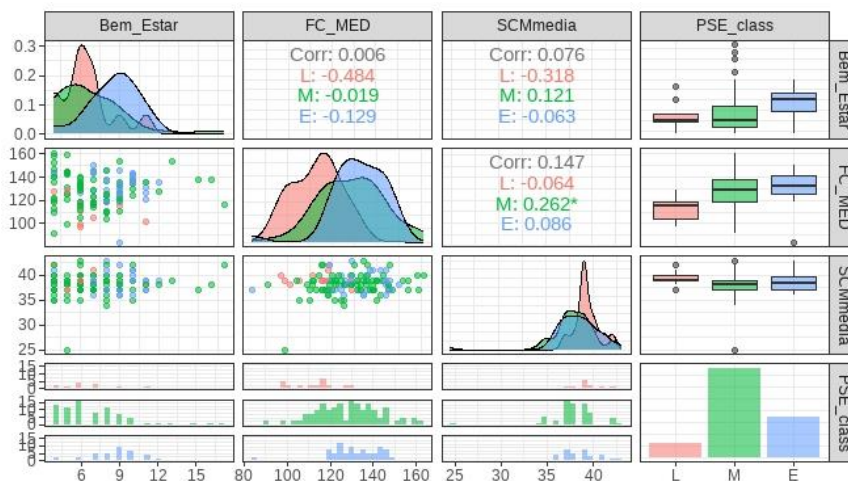
No caso da variável SCMmedia demonstra uma distribuição altamente assimétrica à esquerda. Nesse contexto, a cauda esquerda da distribuição é significativamente mais longa do que a cauda direita. A curtose positiva alta indica que a distribuição é extremamente leptocúrtica, caracterizada por caudas excepcionalmente pesadas e um pico acentuado em relação à distribuição normal. O gráfico 4 a seguir apresenta a distribuição dos dados do SCMmedia.

Gráfico 4. Distribuição dos dados do SCMmedia



O gráfico 5 a seguir apresenta os valores de correlação das variáveis preditoras e suas respectivas distribuições em relação a variável predita em suas diferentes classes. A contagem e distribuição de frequência das classes é de 12 (9,8%) para a classe “Leve”, para a classe “Moderada” é de 76 (62,3%) e, para a classe “Elevada” é de 34 (27,9%). Esses resultados indicam um desbalanceamento acentuado nos dados, principalmente, para a classe “Leve”. Os valores de correlação entre as variáveis Bem_Estar, FC_MED e SCMmedia são fracas e próximas de zero. Portanto, as mudanças em uma dessas variáveis têm uma influência limitada sobre as outras duas. O mesmo pode ser observado quando se analisam as correlações nas diferentes classes.

Gráfico 5. Distribuição e correlação das variáveis preditoras em diferentes classes.



Legenda: FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento; PSE_Class: PSE class: classificação da percepção subjetiva do esforço em Leve (L), Moderada (M) e Elevada (E).

4.2 Árvore de decisão CART

Inicialmente todas as variáveis descritas na tabela 1 foram incluídas no ajuste do modelo. A figura 7 a seguir apresenta o modelo de árvore de decisão.

FC_MED < 138 e Bem_Estar < 9	Moderada	55%
FC_MED ≥ 138 e Bem_Estar < 9	Elevada	70%

Legenda: FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento

Posteriormente o treinamento do modelo foi realizado o teste da qualidade de predição do modelo utilizando teste. O quadro 4 a seguir apresenta a matriz de confusão da árvore de decisão utilizando os dados de teste.

Quadro 4. Matriz de confusão da árvore de decisão

Classes	Observado		
	Leve	Moderada	Elevada
Predito			
Leve	0	0	0
Moderada	13	5	0
Elevada	4	1	0

A tabela 2 a seguir apresenta os valores de sensibilidade, especificidade e acurácia balanceada para as três classes.

Tabela 2. Métricas de avaliação da árvore de decisão para cada classe

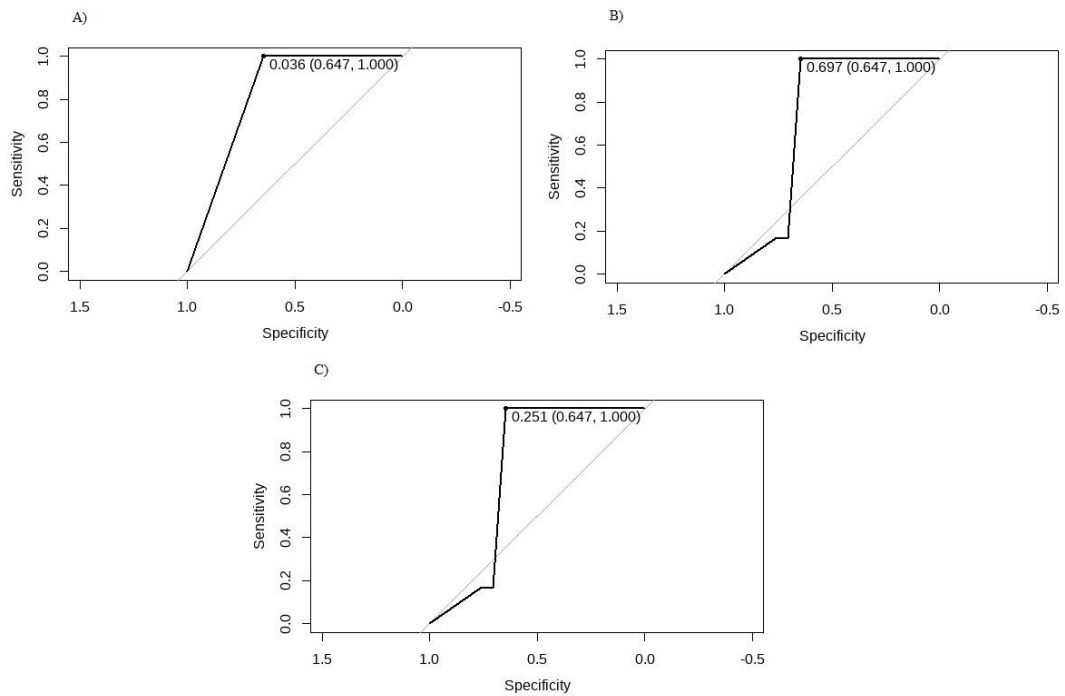
Classe	Sensibilidade	Especificidade	Acurácia Balanceada
Leve	0,0000	1,000	0,5000
Moderada	0,8333	0,2353	0,5343
Elevada	N/A	0,7826	N/A

Legenda: N/A: não se aplica.

A acurácia global do modelo é de 0,2174 e intervalo de confiança de 95% entre 0,0746 a 0,437.

Após realizado o cálculo para área da curva ROC para cada uma das classes separadamente e para as classes em associação, verificou-se uma sensibilidade de 0,647 e especificidade de 1,000 para ponto de corte de 0,036 para a classe Leve com AUC de 0,8235, para a classe Moderada ponto de corte de 0,697 com AUC de 0,7108 e, para a classe Elevada ponto de corte de 0,251 com AUC de 0,7108. Para a AUC multiclasse o valor é de 0,5564. O gráfico 6 a seguir apresenta a curva ROC para cada uma das classes.

Gráfico 6. Curva ROC da árvore de decisão para as diferentes classes



Legenda: A) AUC classe Leve; B) AUC classe Moderada; C) AUC classe Elevada. Sensitivity: Sensibilidade; Specificity: Especificidade.

Um novo teste foi realizado com novo ponto de corte baseado na curva ROC obtendo-se valor de acurácia de 0,2609 (26,09%) com IC95% = 0,1023 (10,23%) – 0,4841 (48,41%). A tabela 3 a seguir apresenta os valores de sensibilidade, especificidade e acurácia balanceada para as três classes.

Tabela 3. Métricas de avaliação da árvore de decisão para cada classe dos dados de teste

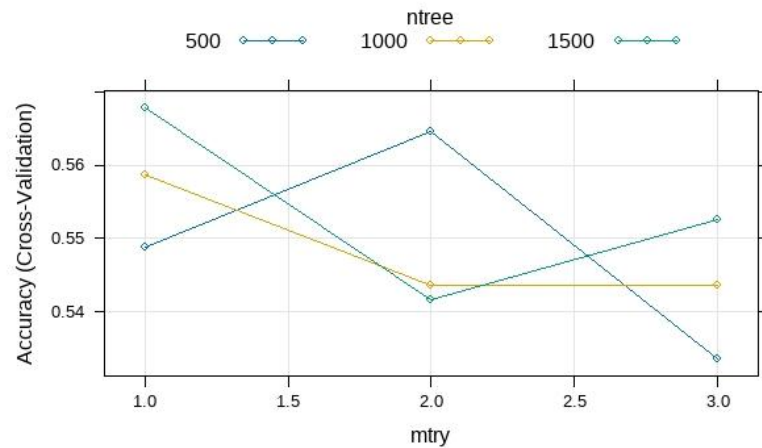
Classe	Sensibilidade	Especificidade	Acurácia Balanceada
Leve	0,0000	1,000	0,5000
Moderada	1,000	0,000	0,5000
Elevada	N/A	1,000	N/A

Legenda: N/A: não se aplica.

4.3 Floresta Aleatória

Para análise da Floresta Aleatória utilizando a validação cruzada com 10 partições, foi encontrada acurácia de 0,5677 e kappa de 0,1055 quando os hiperparâmetros são otimizados utilizando 1 variável treinada a cada nó em 1500 árvores de decisão. O gráfico 7 a seguir apresenta a seleção de hiperparâmetros.

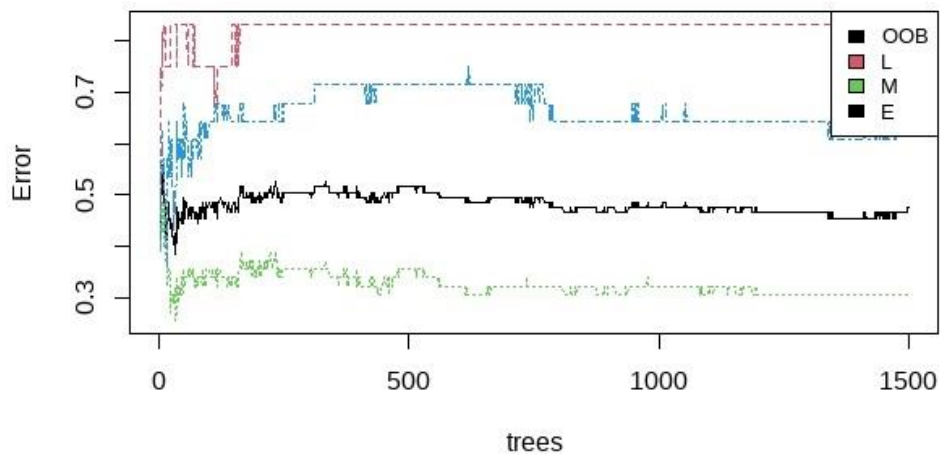
Gráfico 7. Seleção de hiperparâmetros para Floresta Aleatória



Legenda: Accuracy: acurácia; mtry: Tamanho mínimo do nó; ntree: Número de árvores.

O modelo apresentou uma estimativa de erro fora da amostra (*out-of-bag*) de 47,47%. Além disso, a taxa de erro de classificação da PSE Leve é de 0,8333, para a PSE Moderada é de 0,322, enquanto para a PSE Elevada é de 0,6429. O gráfico 8 a seguir ilustra esses resultados.

Gráfico 8. Taxa de erro out of bag da Floresta Aleatória.



Legenda: Error: taxa de erro out-of-bag; trees: número de árvores do modelo. Linha preta OOB: erro geral do modelo; Linha vermelha (L): erro da classe Leve; Linha verde (M): erro da classe Moderada; Linha azul (E): erro da classe Elevada.

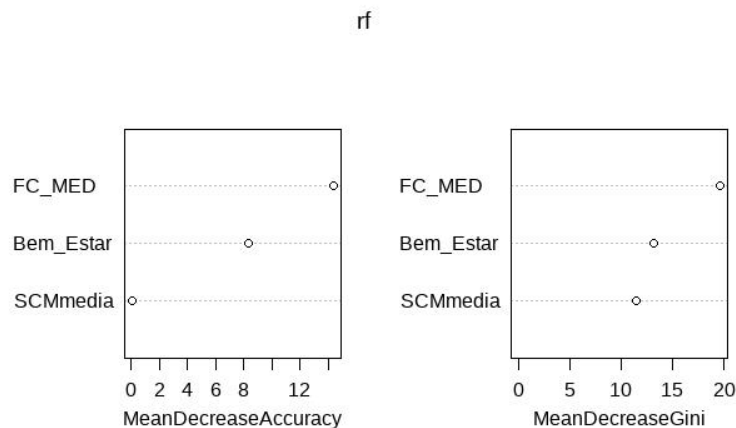
Após aplicado o modelo de Floresta Aleatória nos dados de treinamento obteve-se acurácia de 0,5618 (56,18%) e kappa - 0,017 (- 0,02%). O quadro 5 a seguir apresenta a matriz de confusão para a Floresta Aleatória.

Quadro 5. Matriz de confusão da Floresta Aleatória para os dados de treinamento

Classes	Observado		
	Leve	Moderada	Elevada
Predito			
Leve	2	10	0
Moderada	5	40	14
Elevada	0	18	10

Quando analisadas a média de decréscimo da acurácia por meio de permutação, o Bem_Estar apresenta valor de 8,370, a FC_MED valor de 14,346 e o SCMmedia valor de 0,106. Na avaliação da importância das variáveis utilizando o método Gini a FC_MED exibe o valor de 19,57, o Bem_Estar valor de 13,11 e o SCMmedia valor de 11,40. O gráfico 9 a seguir ilustra esses resultados.

Gráfico 9. Medida de importância das variáveis predictoras no modelo Floresta Aleatória



Legenda: rf: Random Forest; FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento. MeanDecreaseAccuracy: média de decréscimo da acurácia; MeanDecreaseGini: média de decréscimo do índice Gini.

Aplicado aos dados de teste, o modelo de Floresta Aleatória apresentou acurácia de 0,1739 com intervalo de confiança de 95% de 0,0495 a 0,3878. O quadro 6 a seguir apresenta a matriz de confusão dos dados de teste. A Tabela 4 a seguir apresenta os resultados de sensibilidade, especificidade e acurácia balanceada por classe.

Quadro 6. Matriz de confusão da Floresta Aleatória para os dados de teste

Classes	Observado		
	Leve	Moderada	Elevada
Predito			
Leve	0	0	0
Moderada	10	4	0
Elevada	7	2	0

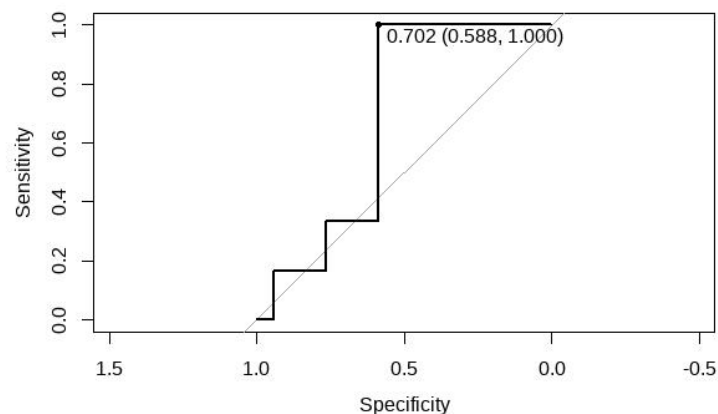
Tabela 4. Métricas de avaliação da Floresta Aleatória para cada classe dos dados de teste

Classe	Sensibilidade	Especificidade	Acurácia Balanceada
Leve	0,0000	1,000	0,5000
Moderada	0,6667	0,4118	0,5392
Elevada	N/A	0,6087	N/A

Legenda: N/A: não se aplica.

Ao realizar o cálculo da AUC com os dados de teste obteve-se valor de 0,6765, sensibilidade geral de 0,588 e especificidade geral igual a 1,000. Ao aplicar o novo ponto de corte (0,702) a acurácia do modelo tendeu a zero. O gráfico 10 a seguir apresenta a curva ROC para o modelo de Floresta Aleatória. A tabela 5 a seguir apresenta os valores de sensibilidade, especificidade e acurácia balanceada por classe.

Gráfico 10. Curva ROC para modelo de Floresta Aleatória



Legenda: Sensitivity: Sensibilidade; Specificity: Especificidade.

Tabela 5. Métricas de avaliação da Floresta Aleatória para cada classe dos dados de teste como novo ponto de corte da curva ROC.

Classe	Sensibilidade	Especificidade	Acurácia Balanceada
Leve	0,0000	1,0000	0,5000
Moderada	0,0000	0,4118	0,2059
Elevada	N/A	0,4348	N/A

Legenda: N/A: não se aplica.

4.3 Bootstrap Aggregating (*Bagging*)

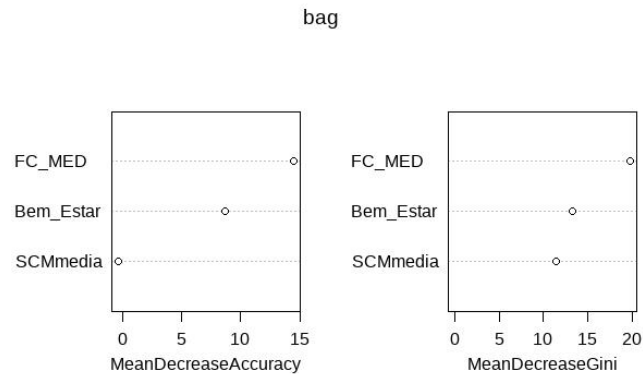
Para análise utilizando *Bagging* foi encontrada acurácia de 0,5217 e kappa de 0,1767 quando os hiperparâmetros são otimizados utilizando 3 variáveis a cada nó em 1500 árvores de decisão. A estimativa de erro fora da amostra (*out-of-bag*) de 48,48%. A taxa de erro de classificação da PSE Leve é de 0,8333, para a PSE Moderada é de 0,3559, enquanto para a PSE Elevada é de 0,6071. O quadro 7 a seguir apresenta a matriz de confusão para o modelo *Bagging*.

Quadro 7. Matriz de confusão do *Bagging* para os dados de teste

Classes	Observado		
	Leve	Moderada	Elevada
Predito			
Leve	2	10	0
Moderada	7	38	14
Elevada	1	16	11

As medidas de importância das variáveis do modelo mantiveram-se similar ao modelo de árvore aleatória. Sendo assim, quando analisadas a média de decréscimo da acurácia por meio de permutação, o Bem_Estar apresenta valor de 8,683, a FC_MED valor de 14,448 e o SCMmedia valor de - 0,307. O gráfico 11 a seguir ilustra esses resultados. Na avaliação da importância das variáveis utilizando o método Gini aparece a FC_MED apresenta valor de 19,780, o Bem_Estar igual a 13,244 e o SCMmedia igual a 11,402. O gráfico 11 a seguir ilustra esses resultados.

Gráfico 11. Medida de importância das variáveis preditoras no modelo Bagging

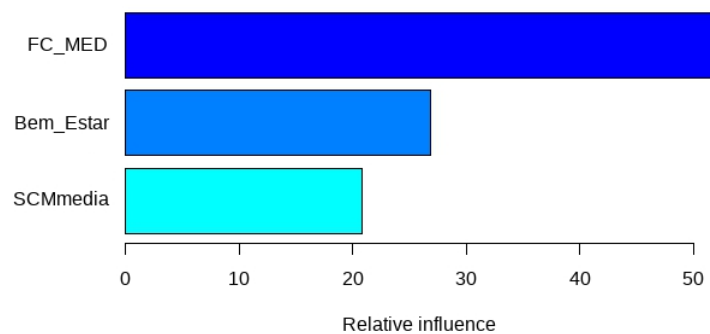


Legenda: bag: Bootstrap Aggreging (Bagging); FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento. MeanDecreaseAccuracy: média de decréscimo da acurácia; MeanDecreaseGini: média de decréscimo do índice Gini.

Ao aplicar os resultados do *Bagging* aos dados de teste, apresentou-se acurácia de 0,5652 (56,52%) e kappa de $-0,0176$ (- 1,76%), valores esses, similares ao modelo de árvores aleatórias.

4.4 Boosting

Para o modelo Boosting foi encontrado para os dados de treinamento acurácia de 0,5346 e kappa de 0,1222 (12,22%). Esses resultados apresentam uma precisão média, o que, embora não seja muito elevada, indica um desempenho acima do acaso. As variáveis preditoras apresentaram valor de importância relativa no modelo de 52,265 para FC_MED, o Bem_Estar apresentou valor de 26,865 e, o SCMmedia valor de 20,869. O gráfico 12 a seguir ilustra esses resultados.

Gráfico 12. Medida de importância das variáveis preditoras no modelo *Boosting*

Legenda: FC_MED: frequência cardíaca média; SCMmedia: média do salto com contramovimento.

A acurácia do modelo aplicado ao conjunto de testes aumentou para 0,6521. Além disso, a estatística Kappa também aumentou para 0,2580.

4.5 Avaliação dos modelos

A avaliação do modelo foi realizada no conjunto de dados de teste resultando nas métricas apresentadas na tabela 6 a seguir.

Tabela 6. Comparação entre modelos.

Métrica	Árvore de Decisão	Floresta Aleatória	<i>Bagging</i>	<i>Boosting</i>
Acurácia	0,2609	0,1739	0,5652	0,6521
Kappa	0,000	0,0200	0,0176	0,2580

5 DISCUSSÃO

O objetivo do presente estudo foi ajustar um modelo preditivo da carga de treinamento de um atleta da seleção brasileira de *Taekwondo* baseado em árvores de decisão de classificação. Dentre os modelos ajustados, o *Boosting* foi o que demonstrou maior valor de acurácia e, portanto, foi mais adequado para prever a PSE quando comparado aos demais modelos.

O monitoramento da PSE é uma prática comum em muitos esportes (Inoue *et al.*, 2022). Normalmente, as informações são usadas para compreender a carga de treinamento e, às vezes, modificar as próximas sessões (Inoue *et al.*, 2022). Identificar a relação de variáveis, fisiológicas, psicológicas e biomecânicas, com a PSE, provavelmente beneficiaria ainda mais a compreensão e gestão da carga de treinamento (Schelling e Robertson, 2020; Montull *et al.*, 2022). Nesse sentido, o modelo proposto pode apresentar aplicabilidade prática e principalmente pode ser um ponto de partida para futuros estudos que buscam utilizar técnicas de *Machine Learning* no contexto do treinamento esportivo.

No presente estudo, ao desenvolver o modelo de árvore de decisão pelo algoritmo CART, a figura 7 e o quadro 3 apresentam as probabilidades de classificação dos diferentes níveis de PSE. Sendo assim, é possível perceber que a aplicação dessa técnica pode reduzir a complexidade no processo de tomada de decisão (James *et al.*, 2013; Rico-González *et al.*, 2023), que envolve escolhas a partir de diferentes opções, que são condicionadas pela quantidade de informação disponível, às vezes limitada ou excessiva, pelas limitações cognitivas do treinador, heurísticas e vieses, pela quantidade finita de tempo disponível e pelos níveis de risco e benefício (Schelling e Robertson, 2020).

No limite do nosso conhecimento, esse é um dos primeiros estudos que objetiva aplicar modelos de *machine learning* para prever a PSE e individualizar a carga de treinamento de um atleta de *Taekwondo*, o que limita nossa discussão. Porém, ao analisar a matriz de confusão da árvore de decisão (quadro 4) verifica-se que, para a PSE “Leve”, nenhuma previsão correta foi obtida, o que reflete uma incapacidade completa do modelo em discernir essa classe. Já para a PSE “Moderada”, embora algumas previsões tenham sido corretas, uma quantidade substancial de classificações incorretas foi observada. Por fim, para a PSE “Elevada” foi encontrada uma situação similar à PSE “Leve”, com todas as previsões incorretas. A acurácia global do modelo foi de 21,74%, considerada baixa, indicando que o modelo raramente acerta as previsões. O intervalo de confiança de 95% variando de 0,07% a 0,44% amplia ainda mais a incerteza associada ao desempenho do modelo.

Quando analisada a sensibilidade para as três classes, a sensibilidade para a PSE “Leve” é nula, sinalizando uma incapacidade completa de identificar verdadeiros positivos nesta classe. Para a PSE “Moderada” a sensibilidade é razoável, mas ainda associada a erros substanciais. Por fim, a PSE “Elevada” não possui verdadeiros positivos, tornando impossível calcular a sensibilidade. Quanto a análise da especificidade, observam-se que as PSE “Leve” e “Elevada” exibem alta especificidade, mas a PSE “Moderada” apresenta especificidade consideravelmente menor. Além disso, as métricas de valores preditivos positivos não podem ser calculadas para algumas classes devido à ausência de verdadeiros positivos.

A acurácia balanceada de 50% sugere que o modelo está tomando decisões igualmente balanceadas em relação à PSE “Leve”, o que pode não ser útil na prática, dada a falta de previsões corretas para esta classe. Para a PSE “Moderada”, o valor de acurácia balanceada está um pouco acima de 0,5, indicando que o modelo tem uma capacidade moderada de prever esta classe. No entanto, ainda está longe de ser um desempenho sólido, considerando que a acurácia balanceada varia de 0,0 (nenhuma previsão correta) a 1,0 (todas as previsões corretas). Por fim, não foi possível calcular a acurácia balanceada para a PSE “Elevada”, uma vez que não há verdadeiros positivos para esta classe. Isso reflete a incapacidade do modelo em prever corretamente esta classe.

Uma possível explicação para esse baixo desempenho do modelo de árvore de decisão de classificação, pode estar relacionada a falta de consenso na literatura do que corresponde a PSE “Leve”, “Moderada” ou “Elevada”, a partir das diferentes escalas, CR-10, Borg (6-20) como apresentada por Inoue e colaboradores (Inoue *et al.*, 2022). Portanto, de acordo com o referencial utilizado, a quantidade de sessões classificadas como PSE “Leve” pode aumentar e/ou diminuir, aumentando assim, o balanceamento ou desbalanceamento entre as diferentes classes. Além disso, outros estudos demonstram que é moderada a correlação entre a PSE intencionada pelo treinador ao prescrever a carga de treinamento e, a PSE observada durante a sessão e, a PSE reportada pelo atleta ao final da sessão (Paul *et al.*, 2021). Isso ocorre devido aos inúmeros fatores intervenientes relacionados a carga de treinamento e que muitas das vezes não são possíveis de serem controlados durante o processo (Haddad *et al.*, 2017; Paul *et al.*, 2021). Portanto, outras variáveis, desconhecidas, não controladas durante a coleta de dados podem ser fator confusão que interferiram na classificação da PSE pelo atleta (Haddad *et al.*, 2017; Paul *et al.*, 2021) e, conseqüentemente na capacidade preditiva do modelo.

No novo teste, realizado a partir do novo ponto de corte baseado na curva ROC, obteve-se valor de acurácia de 26,09% com IC95% de 10,23% a 48,41%. Sendo assim, é possível perceber que não houve melhora substancial em relação ao primeiro modelo, o que indica que o modelo não consegue identificar de forma eficaz os casos verdadeiramente positivos para a PSE “Leve”. Porém, como verificado na tabela 3, o modelo possui uma alta sensibilidade para a PSE “Moderada”, o que significa que ele é capaz de identificar corretamente os casos dessa classe. Quanto a especificidade, os resultados sugerem que o modelo é altamente específico para a PSE “Leve”, mas não consegue identificar corretamente os verdadeiros negativos da PSE “Moderada”. Para a PSE “Elevada”, a especificidade indica um alto nível. A acurácia balanceada possui um desempenho moderado em termos de precisão global, apesar das variações nas métricas de sensibilidade e especificidade em diferentes classes. Sendo assim, interpretação dos resultados sugere que o modelo é mais eficaz na identificação de casos da PSE “Moderada”, enquanto sua capacidade nas outras classes é mais baixa, especialmente na PSE “Leve”.

Esses resultados corroboram com estudos de associação que verificaram que treinadores, mesmo os mais experientes, têm dificuldade de prever níveis de PSE mais baixos, correspondentes a PSE “Leve” (Paul *et al.*, 2021; Inoue *et al.*, 2022). Porém, são mais assertivos ao preverem níveis de PSE “Moderada”, seguido dos níveis de PSE “Elevada” (Paul *et al.*, 2021; Inoue *et al.*, 2022). Uma possível explicação para a dificuldade de predição de PSE “Leve” e “Elevada”, tanto pelos treinadores (Paul *et al.*, 2021; Inoue *et al.*, 2022) quanto pelo modelo de árvore de decisão de classificação do presente estudo, pode ser pela menor quantidade de sessões de treinamento prescritas na PSE “Leve”. Sendo assim, o atleta teria maior dificuldade de perceber o que é um treinamento de PSE “Leve” quando esse ocorre, justamente por ter vivenciado menos essa classe. Além disso, há uma tendência da carga de treinamento a regredir para a média em vez de permanecer polarizada (Inoue *et al.*, 2022), o que pode acarretar em mais sessões de treinamento classificadas com PSE “Moderada”, mesmo quando essas são “Leves” e/ou “Elevadas”, o que explicaria a melhor capacidade de predição do modelo para esta classe.

Ao executar a técnica de Floresta Aleatória, a acurácia aumentou para 56,77%, entretanto, o modelo apresentou uma estimativa de erro fora da amostra (*out-of-bag*) de 47,47%, o que indica um desempenho insatisfatório na tarefa de classificação. Além disso, a taxa de erro para classificação foi de 83,33% para PSE “Leve”, para PSE “Moderada” o erro foi de 32,20% e, de 64,29% para PSE “Elevada”. Nos dados de treinamento a acurácia do modelo foi de 56,18%.

Avaliando o decréscimo da acurácia pelo método de permutação verificou-se que a FC_MED exhibe a maior relevância, demonstrando um impacto substancial na acurácia do modelo quando sujeita a permutações aleatórias. Por outro lado, a variável Bem_Estar também contribui para o desempenho do modelo, embora seu efeito seja menos pronunciado do que em comparação com FC_MED. No entanto, a variável SCMmedia exhibe uma importância relativamente baixa, indicando que sua influência sobre a capacidade preditiva do modelo é nula. Na avaliação da importância das variáveis utilizando o método Gini, a FC_MED exhibe o maior valor, indicando que sua inclusão é fundamental para a redução da impureza nos nós de decisão. O Bem_Estar continua em segundo lugar e por último o SCMmedia. Esses resultados demonstram que FC_MED é a variável mais importante na tomada de decisões do modelo, seguida pelo Bem_Estar, enquanto SCMmedia desempenha um papel menos influente, mas ainda relevante em menor grau.

Apesar da PSE estar intimamente ligada aos sinais neurais relacionados a magnitude do comando motor (Monjo e Allen, 2023), esta apresenta correlação moderada a forte com a frequência cardíaca (Borg, 1982; Haddad *et al.*, 2017), o que também foi verificado no presente estudo. Assim, justificam-se os valores mais elevados de FC_MED na avaliação de importância das variáveis preditoras. Além disso, também é reportado na literatura que fatores biopsicossociais estão relacionados a como o atleta percebe o esforço de treinamento (Haddad *et al.*, 2017). Sendo assim, como a escala de Bem_Estar leva em considerações métricas de estresse, fadiga, dor e sono para seu cálculo, esses fatores podem justificar o nível de importância dessa variável na predição da PSE. Entretanto, esse resultado deve ser analisado com cautela, pois estudo anterior demonstrou que os itens de Bem_Estar têm capacidade preditiva limitada em relação as medidas de carga interna e externa (Campbell *et al.*, 2021). Por sua vez, apesar da altura do SCMmedia ser uma ferramenta bastante utilizada para o controle da carga de treinamento no *Taekwondo* (Silva *et al.*, 2023), esta é mais eficaz para detectar alterações no sistema neuromuscular (Claudino *et al.*, 2017) e, pode não ser sensível suficiente para oferecer informações relevantes para predizer a carga interna medida por meio da PSE.

Aplicado aos dados de teste, o modelo de Floresta Aleatória apresentou acurácia de apenas 17,39% com intervalo de confiança de 95% variando de 4,95% a 38,78%. A análise das estatísticas por classe revela que para a PSE “Leve”, a sensibilidade registrada indica que o modelo não conseguiu identificar corretamente nenhuma instância dessa classe. Por outro lado, para a PSE “Moderada”, a sensibilidade sugere uma capacidade relativamente melhor de identificar corretamente as instâncias dessa classe. Quanto à PSE “Elevada”, a sensibilidade

reportada indica que essa classe não foi corretamente reconhecida pelo modelo. Além disso, a especificidade varia muito entre as classes. A acurácia balanceada apresenta maior valor para a PSE “Moderada”, enquanto para a PSE “Leve” o valor é menor. Para a PSE “Elevada” não foi possível calcular a acurácia balanceada, o que indica necessidade de aprimoramento do modelo para atingir equilíbrio de sensibilidade e especificidade em todas as classes. Ao realizar o cálculo da AUC com os dados de teste e, ao aplicar o novo ponto de corte a acurácia do modelo tendeu a zero, conseqüentemente, os valores de sensibilidade, especificidade e acurácia balanceada apresentaram redução, como evidenciado na Tabela 5.

Na análise do modelo *Bagging* foi verificada acurácia de 52,17% e, assim como o modelo de árvores aleatórias, o modelo *Bagging* apresentou uma estimativa de erro fora da amostra (*out-of-bag*) elevado. Além disso, a taxa de erro de classificação da PSE “Leve” é de 83,33%, para a PSE “Moderada” é de 35,59%, enquanto para a PSE “Elevada” é de 60,71%. As medidas de importância das variáveis do modelo mantiveram-se similar ao modelo de Floresta Aleatória. Sendo assim, quando analisadas a média de decréscimo da acurácia por meio de permutação, a FC_MED é seguida pelo Bem_Estar e por último o SCMmedia. Na avaliação da importância das variáveis utilizando o método Gini, manteve-se a FC_MED em primeiro lugar, seguida pelo Bem_Estar e por último o SCMmedia. Ao aplicar os resultados do *Bagging* aos dados de teste, verificou-se acurácia de 56,52%, valor bem superior ao da Floresta Aleatória (17,39%) e árvore de decisão (26,09%).

Os resultados observados para cada modelo podem ser explicados pela capacidade de cada técnica de modelagem em lidar com o *overfitting* e a generalização (Hastie, Tibshirani e Friedman, 2009; Kassambara, 2017). O *Bagging* é uma técnica que combina múltiplos modelos de *machine learning* para melhorar o desempenho geral. Ela envolve treinar vários modelos independentes em subconjuntos diferentes dos dados de treinamento e, em seguida, combinar suas previsões (Hastie, Tibshirani e Friedman, 2009; Kassambara, 2017). Isso pode resultar em um modelo geral mais robusto e preciso. Por outro lado, uma Floresta Aleatória pode ser suscetível ao *overfitting*, especialmente se o número de árvores na floresta for muito alto ou se os hiperparâmetros não forem ajustados corretamente (Hastie, Tibshirani e Friedman, 2009; Kassambara, 2017).

Quando utilizada a técnica de *Boosting*, foi encontrado para os dados de treinamento acurácia de 53,46%. Esses resultados apresentam uma precisão média, o que, embora não seja muito, indica um desempenho acima do acaso. As variáveis preditoras mantiveram sua ordem

importância relativa no modelo, na qual a FC_MED aparece em primeiro lugar, o Bem_Estar aparece em segundo lugar e o SCMmedia apresenta menor grau de importância em último lugar. A análise dos resultados após a aplicação do modelo aos dados de teste revela melhorias significativas em relação ao desempenho anterior. A acurácia do modelo aumentou, indicando que o modelo classificou corretamente aproximadamente 65,21% das classes no conjunto de teste. Além disso, a estatística Kappa também aumentou, o que indica uma concordância 25,80% melhor do que o esperado ao acaso.

Ao comparar os modelos na Tabela 6, podemos perceber que o modelo de Floresta Aleatória foi o que apresentou pior desempenho em classificar o nível de PSE quando comparado aos demais modelos. Uma possível explicação para esse baixo desempenho pode estar relacionada ao desequilíbrio de classes, ou seja, quando uma classe é muito mais prevalente do que as outras, a Floresta Aleatória pode ser menos eficaz na classificação da classe minoritária (Kassambara, 2017; Izbicki e Santos, 2020). Por outro lado, o modelo *Boosting* apresentou o maior aumento tanto na acurácia quanto no valor Kappa, o que indica desempenho melhor em relação aos outros modelos. Isso se dá pelo fato que o *Boosting* é mais eficaz em lidar com conjuntos de dados desbalanceados, onde algumas classes possuem muito menos exemplos do que outras (Kassambara, 2017; Izbicki e Santos, 2020). Sendo assim, o *Boosting* atribui mais importância às instâncias das classes minoritárias, ajudando a evitar o viés em direção às classes majoritárias que outros modelos podem apresentar (Kassambara, 2017; Izbicki e Santos, 2020). Portanto, o *Boosting* é uma técnica que se destaca na correção de erros, na adaptação a dados desbalanceados e na redução da variância do modelo. Sua capacidade de combinar modelos fracos de forma adaptativa e sua ênfase nas instâncias difíceis de classificar contribuem para seu melhor desempenho em relação a outros modelos (Kassambara, 2017; Izbicki e Santos, 2020).

Com base nos resultados do presente estudo, sugere-se cautela nas análises de predição da carga de treinamento, haja vista a elevada variabilidade nas respostas de FC_MED, SCMmedia e Bem_Estar em relação a PSE nas diferentes classes, o que pode resultar na baixa capacidade preditiva dos diferentes modelos. Portanto, estudos futuros devem levar em consideração alguns fatores que podem ter influenciado na construção dos modelos, como por exemplo, o número de sessões de treinamento analisadas e, o controle de outras variáveis biopsicossociais que podem interferir nas PSE (Haddad *et al.*, 2017). Além disso, a característica da modalidade, se individual ou coletiva, o nível competitivo do atleta, seja ele recreacional, regional, nacional e/ou internacional e, a fase de periodização do treinamento, também podem influenciar na PSE

(Paul *et al.*, 2021; Inoue *et al.*, 2022) e talvez também na capacidade preditiva dos modelos. Entretanto, essas limitações parecem não invalidar a aplicabilidade de técnicas de *machine learning* para auxílio no planejamento, monitoramento e controle da carga de treinamento, já verificada em estudos anteriores (Davidson *et al.*, 2020; Campbell *et al.*, 2021; Leeuw *et al.*, 2022; Biró *et al.*, 2023; Rico-González *et al.*, 2023).

6 CONCLUSÃO

Modelos de *Machine Learning* que utilizam algoritmos de árvore de decisão de classificação podem ser aplicados na predição da carga de treinamento para auxiliar no planejamento, monitoramento e controle da carga de treinamento, contudo, com os dados utilizados no presente estudo, o método *Bossting* se mostrou o mais adequado para desempenhar tal função, mas as predições devem ser analisadas com cautela.

REFERÊNCIAS

- ALBUQUERQUE, M. R., MESQUITA, P. H. C., HERRERA-VALENZUELA, T., FRANCHINI, E. Predicting taekwondo winners in high Level competition using ranking scores and country performance scores: an analysis of the 2019 world taekwondo championship. **Journal of Martial Arts Anthropology**. v. 21, n. 2, p. 19-26, 2021.
- ANDREATO, L. V., LARA, F. J. D., ANDRADE, A. e BRANCO, B. H. M. Physical and physiological profiles of Brazilian jiu-jitsu athletes: a systematic review. **Sports Medicine**. v. 3, n. 9, p. 2-17, 2017.
- Arel-Bundock V. Modelsummary: Data and Model Summaries in R. **Journal of Statistical Software**, v. 103, n. 1, p. 1–23, 2022.
- ARMSTRONG, L., BERGERON, M. F., LEE, E. C., MERSHON, J. E., ARMSTRONG, E. M. Overtraining syndrome as a complex systems phenomenon. **Frontiers in Network Physiology**. v. 18, n. 1, p. 1-20, 2022.
- BIRÓ, A., SZILÁGYI, S. M., SZILÁGYI, L., MARTÍN-MARTÍN, J., CUESTA-VARGAS, A. I. *Machine Learning* on prediction of relative physical activity intensity using medical radar sensor and 3d accelerometer. **Sensor**. v. 23, n. 7, p. 3595-3623, 2023.
- BORG, G. A. V. Psychological bases of perceived exertion. **Medicine and Science in Sports and Exercise**. v. 14, n. 5, p. 377-381, 1982.
- BORRESEN, J. e LAMBERT, M. I. Autonomic control of heart rate during and after exercise: measurements and implications for monitoring training status. **Sports Medicine**. v. 38, n. 8, p. 633-646, 2008.
- BORRESEN, J. e LAMBERT, M. I. Quantifying training load: A comparison of subjective and objective methods. **International Journal of Sports Physiology and Performance**. v. 3, p. 16–30, 2008b.
- BORRESEN, J. e LAMBERT, M.I. The quantification of training load, the training response and the effect on performance. **Sports Medicine**. v. 39, n. 9, p. 779-795, 2009.
- BOURDON *et al.* Monitoring athlete training loads: Consensus Statement. **International Journal of Sports Physiology and Performance**. v. 12, p. 161-170, 2017.
- BREIMAN, L. *et al.* **Classification And Regression Trees**. Routledge, 1984.
- CAMPBELL, B.I., BOVE, D., WARD, P., VARGAS, A. e DOLAN, J. Quantification of training load and training response for improving athletic performance. **Strength and Conditioning Journal**. v. 39, n. 5, p. 3-13, 2017.
- CAMPBELL, P.G., STEWART, I.B., SIROTIC, A.C., DROVANDI, C., FOY, B.H., MINETT, G.M. Analysing the predictive capacity and dose-response of wellness in load monitoring. *Journal of Sports Science*. v. 39, n. 12, p. 1339-1347, 2021.
- CHAUDHARY, S., YADAV, S., KUSHWAHA, S., SHAHI, S. R. P. A brief review if Machine Learning and its application. **SAMRIDDHI a Journal of Physical Sciences, Engineering and technology**. v. 12, n. 1, 2020.

- CHEN, T., e GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD **International Conference on Knowledge Discovery and Data Mining**. p. 785–794, 2016.
- CLAUDINO, J.G *et al.* The countermovement jump to monitor neuromuscular status: A meta-analysis. **Journal of Science and Medicine in Sport**. v. 20, p. 397-402, 2017.
- DAVIDSON, P., DÜKING, P., ZINNER, C., SPERLICH, B., HOTH, A. Smartwatch Derived Data and Machine Learning Algorithms Estimate Classes of Ratings of Perceived Exertion in Runners: A Pilot Study. **Sensors**. v. 20, n. 9, p. 2637, 2020.
- DIDORF, C., BARTAGUIZ, E., GASSMAN, F., FRÖHLICH, M. Conceptual structure and current trends in artificial intelligence, Machine Learning, and deep learning research in sports: a bibliometric review. **International Journal of Environmental Research and Public Health**. v. 20, n. 173, p. 1-23, 2023.
- FÁVERO, L. P., BELFIORE, P. e SOUZA, R. F. **Data Science, Analytics and Machine Learning with R**. Academic Press. Elsevier. 2023.
- FERREIRA, E. B. e OLIVEIRA, M. S. **Introdução à Estatística com R**. Editora Universidade Federal de Alenas, 2020.
- FOSTER, C. *et al.* A new approach to monitoring exercise training, **The Journal of Strength and Conditioning Research**. v. 15, n. 1, p. 109-115, 2001.
- FOSTER, C., MARROYO, A.R. e KONING, J.J. Monitoring training loads: the past, present, and the future. **International Journal of Sports Physiology and Performance**. v. 12, n. 2. p. 2-8, 2017.
- FOX, J. L., STANTON, R., SARGENT, C., WINTOUR, S., SCANLAN, A. T. The association between training load and performance in team sports: a systematic review. **Sports Medicine**. v. 48, n. 12, p. 2743-2774, 2018.
- FRANCHINI, E., DEL VECCHIO, F. B., MATSUSHIGUE, K. A. e ARTIOLI, G. G. Physiological profile of elite judo athletes. **Sports Medicine**. v. 41, n. 2, p. 147-166, 2011.
- FURRER, R., HAWLEY, J. A., HANDSHIN, C. The molecular athlete: exercise physiology from mechanisms to medals. **Physiological Reviews**. v. 103, p. 1693-1787, 2023.
- GABBETT, T. The training injury prevention paradox: should athletes be training smarter and harder? **British Journal of Sports Medicine**. v. 50, p. 273-280, 2016.
- GATHERCOLE, R., SPORER, B., STELLINGWERFF, T. e SLEIVERT, G. Alternative countermovement jump analysis to quantify acute neuromuscular fatigue. **International Journal of Sports Physiology and Performance**. v. 10, p. 84-92, 2015.
- GHELLER, R.G., DAL PUPO, J., LIMA, L.A.P., MOURA, B.M. e SANTOS, S.G. A influência da profundidade de agachamento no desempenho e em parâmetros biomecânicos do salto com contramovimento. **Revista Brasileira de Cineantropometria e Desempenho Humano**. v. 16, n. 6, p. 658-668, 2014.
- GOULART *et al.* Correlation between roundhouse kick and countermovement jump performance. **Archives of Budo**. v. 12, p. 125-131, 2016.
- GREENWELL, B., BOEHMKE, B., CUNNINGHAM, J. GBM Developers. gbm: Generalized Boosted Regression Models. **R package** version 2.1.8. 2020.

- HADDAD, M., STYLIANIDES, G., DJAOUL, L., DELLAL, A. e CHAMARI, K. Session-RPE method for training load monitoring: validity, ecological usefulness, and influencing factors. **Frontiers in Neuroscience**. v. 11, p. 1-14, novembro, 2017.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. New York: Springer, second edition, 2009.
- HELWIG *et al.* Relationship between external, wearable sensor based an internal parameter: a systematic review. **Sensors**. v. 23, n. 2, p. 827-849, 2023.
- HOOPER, S. L. e MACKINNON, L. T. Monitoring overtraining in athletes' recommendations. **Sports Medicine**. v. 20, n. 5, p. 321-327, 1995.
- INOUE, A., BUNN, P. S., CARMO, E. C., LATTARI, E. e SILVA, E. B. Internal training load perceived by athletes and planned by coaches: a systematic review and meta-analysis. **Sports Medicine Open**. v. 8, n. 35, p. 1-32, 2022.
- IZBICKI, R e SANTOS, T.M. **Aprendizado de máquina: uma abordagem estatística**. Livro Eletrônico. São Paulo. ISBN: 978-65-00-02410-4.
- JAMES, G. *et al.* **An Introduction to Statistical Learning**. Springer New York, 2013.
- KASSAMBARA, A. **Machine Learning Essentials - Practical Guide in R**. Birmingham: Packt Publishing, 2017.
- KAUTZNER JR, N. M. Estado da arte das escalas de percepção subjetiva do esforço. **Revista Brasileira de Prescrição e Fisiologia do Exercício**. v. 7, n. 39, p. 293-308, 2013.
- KIELY, J. Periodization theory: confronting an inconvenient truth. **Sports Medicine**. v. 48, n. 4, p. 753-764, 2018.
- KUHN, M., JOHNSON, K. **Applied Predictive Modeling**. Springer New York, 2013
- KUHN, M. Building Predictive Models in R Using the caret Package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008.
- LAURENT *et al.* A practical approach to monitoring recovery: development of perceived recovery status scale. **Journal Strength and Conditioning Research**. v. 25, n. 3, p. 620-628, 2011.
- LEEuw, A. W., VAN BARR, R., KNOBBE, A., ZWAARD, S. Modeling match performance in elite volleyball players: importance of jump load and strength training characteristics. **Sensors**. v. 22, n. 20, p. 7996-7809, 2022.
- LOTURCO, I. e NAKAMURA, F. Y. Training periodisation an obsolete methodology? **Sports Medicine Journal**. p. 110-115, 2016.
- MILBORROW, S. rpart.plot: plot “rpart” models: an enhanced version of “plot.rpart”, **R package** version 3.0.8. 2019.
- McGUIGAN, H. E., HASSMÉN, P., ROSIC, N., STEVENS, C. J. Monitoring training in high performance athletes: what practitioners do? **The Journal of Sport Exercise Science**. v. 5, n. 2, p. 121-129, 2021.
- MONJO, F. e ALLEN, T. What if muscle spindles were also involved in the sense of effort? **The Journal of Physiology**. v. 601, n. 16, p. 3453-3459, 2023.

- MONTULL, L., DACKEVICIENE, A. S., KIELY, J., HRISTOVSKI, R., BALAGUÉ, N. Integrative proposals of sports monitoring: subjective outperforms objective monitoring. *Sports Medicine*. v. 8, n. 42, p. 1-10, 2022.
- MUJIKA, I., HALSON, S., BURKE, L.M., BALAGUÉ, G., FARROW, D. An integrated, multifactorial approach to periodization for optimal performance in individual and team sports. ***International Journal of Sports Physiology and Performance***. v. 13, n. 5, p. 538-561, 2018.
- OTA, S. e KIMURA, M. Statistical injury prediction for professional sumo wrestlers: Modeling and perspectives. ***Plos One***. 2023.
- OUERGUI *et al.* Development of noncontact kickboxing circuit training protocol that simulates elite male kickboxing competition. ***Journal of Strength and Conditioning Research***. v. 29, n. 12, p. 3405-3411, 2015.
- PAUL, D., READ, P., FAROOQ, A. e JONES, L. Factors influencing the association between coach and athlete rating of exertion: a systematic review and meta-analysis. ***Sports Medicine***. v. 7, n. 1, p. 1-20, 2021.
- RAHLF *et al.* A Machine Learning approach to identify risk factors for running related injuries: study protocol for prospective longitudinal cohort trial. ***BMC Sport Science, Medicine and Rehabilitation***. v. 14, n. 75, p.1-11, 2022.
- RICO-GONZÁLEZ, M., PINO-ORTEGA, J., MÉNDEZ, A., CLEMENTE, F. M., BACA, A. Machine Learning application in soccer: a systematic review. ***Biology of Sports***. v. 24, n. 1, p. 249-263, 2023.
- ROMMERS *et al.* A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. ***Medicine and Science in Sports and Exercise***. v. 58, n. 8, p. 1745-1751, 2020.
- ROSSI, A., PAPPALARDO, L., CINTIA, P., LAIA, F. M., FERNÁNDEZ, J., MEDINA, D. Effective injury forecasting in soccer with GPS training data machine learning. ***PlosOne***. v. 13, n. 7, e02011264, 2018.
- SANTOS, J. F. S., WILSON, V. D., HERRERA-VALENZUELA, T. e MACHADO, F. S. M. Time-motion analysis and physiological responses to Taekwondo combat on juvenile and adult athletes: a systematic review. ***Strength and Conditioning Journal***. v. 42, n. 2, p.103-121, 2020.
- SCHELLING, X. e ROBERTSON, S. A development framework for decision support systems in high performance sport. ***International Journal of Computer Science in Sport***. v. 19, n. 1, p. 1-23, 2020.
- SILVA *et al.* Effect of 12 hour-fasting promoted by breakfast omission on acute weight loss and physical performance of Taekwondo athletes. ***Archives of Budo Science of Martial Arts and Extreme Sports***. v. 16, p. 29-35, 2020.
- SILVA *et al.* Intermittent fasting promotes weight loss without decreasing performance in Taekwondo. ***Nutrients***. v.15, p. 3131, 2023.
- SOLIGARD *et al.* How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk injury. ***British Journal of Sports Medicine***. v. 50, n. 17, p. 1030-1041, 2016.

- STAGNO, K. M., THATCHER, R., Van SOMEREN, K. A. A modified TRIMP to quantify the in-season training load of team sports players. **Journal of Sports Science**.v. 25, n. 6, p. 629-634, 2007.
- SZMUCHROWSKI, L. A. e COUTO, B. P. **Treinamento Esportivo: Sistema Integrado do Treinamento**, editora manole ltda, p. 3-26, 2013.
- SZMUCHROWSKI, L. A., CLAUDINO, J. G. O; NETO, S. L. A; MENZEL, H. K; COUTO, B. P. Determinação do número mínimo de saltos verticais para monitorar as respostas ao treinamento pliométrico. **Motricidade**, v. 8, n. 2, p. 383-392, 2012.
- TABORRI, J., MOLINARO, L., SANTOSPAGNUOLO, A., VETRANO, M., VULPIANI, M. C., ROSSI, S. A machine learning approach to measure the anterior cruciate ligament injury risk in female basketball players. **Sensors**. v. 21, p. 3141-3158, 2021.
- THEOPHANO MITSA. **Temporal Data Mining**. Chapman & Hall. Data Mining and Knowledge Discovery Series. 2010.
- THERNEAU, T., ATKINSON, B. e RIPLEY, B. Rpart: Recursive Partitioning. **R Package** Version 4.1-3. 2013.
- VAN DEN TILLAAR, R., BHANDURGE, S., STEWART, T. Can machine learning with IMUs be used to detect different throws and estimate ball velocity in team handball? **Sensors**. v. 21, n. 7, p. 2107-2288, 2021
- VANRENTERGHEM, J., NEDERGAARD, N. J., ROBINSON, M.A. e DRUST, B. Training load monitoring team sports: a novel framework separating physiological biomechanical load adaptation pathways. **Sports Medicine**. v. 47, n. 11, p. 2135-2142, 2017.
- WEST *et al.* More than a metric: how training load is used in elite sport for athlete management. **International Journal of Sports Medicine**. v. 42, n. 4, p. 300-306, 2020.