USO DE MACHINE LEARNING ATRAVÉS DE COMITÊS DE ALGORITMOS E MODELOS ISOLADOS PARA CLASSIFICAÇÃO DE ATLETAS DO GÊNERO MASCULINO EM PROVAS DE IRONMAN 70.3 EM NÍVEIS “APENAS CONCLUINTE” E “COMPETITIVO”

USE OF MACHINE LEARNING THROUGH ALGORITHM COMMITTEES AND ISOLATED MODELS FOR THE CLASSIFICATION OF MALE ATHLETES IN IRONMAN 70.3 RACES INTO "FINISHER" AND "COMPETITIVE" LEVELS

JOSÉ LUCAS SILVA GALDINO
UNIDADE DE POS GRADUAÇÃO, EXTENSÃO E PESQUISA - CENTRO PAULA SOUZA

NAPOLEÃO VERARDI GALEGALE
CENTRO PAULA SOUZA

JULIANE BORSATO BECKEDORFF PINTO
UNIDADE DE POS GRADUAÇÃO, EXTENSÃO E PESQUISA - CENTRO PAULA SOUZA

BRUNA DOS SANTOS SAMPAIO

Comunicação:
O XIII SINGEP foi realizado em conjunto com a 13th Conferência Internacional do CIK (CYRUS Institute of Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

USO DE MACHINE LEARNING ATRAVÉS DE COMITÊS DE ALGORITMOS E MODELOS ISOLADOS PARA CLASSIFICAÇÃO DE ATLETAS DO GÊNERO MASCULINO EM PROVAS DE IRONMAN 70.3 EM NÍVEIS “APENAS CONCLUINTE” E “COMPETITIVO”

Objetivo do estudo
O objetivo do estudo foi comparar o desempenho de quatro algoritmos de Machine Learning, isolados e em comitês (ensembles), para classificar atletas masculinos de IRONMAN 70.3. A pesquisa buscou determinar a eficácia dos modelos na distinção entre níveis "competitivo" e "apenas concluinte".

Relevância/originalidade
A relevância está na aplicação de comitês de algoritmos (ensembles) como um mecanismo inovador para a análise de desempenho no triatlo. A originalidade reside na comparação direta entre modelos isolados e comitês para classificar atletas, explorando uma abordagem precisa.

Metodologia/abordagem
A metodologia consistiu em uma pesquisa exploratória e empírica utilizando uma base de dados do Kaggle. Foram treinados e comparados quatro algoritmos de Machine Learning e comitês (ensembles) para classificar o desempenho de atletas, usando tempos parciais e dados etários como variáveis.

Principais resultados
Os principais resultados indicaram que a diferença de desempenho entre os modelos isolados e os comitês foi pequena. Isso sugere que os modelos individuais são viáveis para essa tarefa, enquanto os comitês podem ser uma ferramenta útil para cenários mais complexos.

Contribuições teóricas/metodológicas
A principal contribuição metodológica foi a aplicação e comparação de comitês de hard e soft voting em um domínio esportivo específico. Teoricamente, o estudo contribui ao demonstrar que, para este problema, modelos mais simples podem ser tão eficazes quanto ensembles complexos.

Contribuições sociais/para a gestão
As principais contribuições são a possibilidade de criação de uma ferramenta que pode auxiliar treinadores e atletas a otimizar treinamentos e estratégias de prova. Socialmente, o estudo promove o uso de dados para uma análise mais objetiva e criteriosa do desempenho atlético.

Palavras-chave: Aprendizagem de Máquina, Comitê, IRONMAN 70.3, Classificação de Atletas, Triatlo

USE OF MACHINE LEARNING THROUGH ALGORITHM COMMITTEES AND ISOLATED MODELS FOR THE CLASSIFICATION OF MALE ATHLETES IN IRONMAN 70.3 RACES INTO "FINISHER" AND "COMPETITIVE" LEVELS

Study purpose
The objective of the study was to compare the performance of four Machine Learning algorithms, individually and in ensembles, in classifying male IRONMAN 70.3 athletes. The research sought to determine the effectiveness of the models in distinguishing between "competitive" and "completer" levels.

Relevance / originality
The relevance lies in the application of algorithmic committees (ensembles) as an innovative mechanism for performance analysis in triathlon. The originality lies in the direct comparison between isolated models and committees for classifying athletes, exploring a precise approach.

Methodology / approach
The methodology consisted of exploratory and empirical research using a Kaggle database. Four machine learning algorithms and ensembles were trained and compared to classify athletes' performance, using split times and age data as variables.

Main results
The main results indicated that the performance difference between the stand-alone models and the ensembles was small. This suggests that stand-alone models are viable for this task, while ensembles may be a useful tool for more complex scenarios.

Theoretical / methodological contributions
The main methodological contribution was the application and comparison of hard and soft voting committees in a specific sports domain. Theoretically, the study contributes by demonstrating that, for this problem, simpler models can be as effective as complex ensembles.

Social / management contributions
The main contributions include the possibility of the creation of a tool that can help coaches and athletes optimize training and competition strategies. Socially, the study promotes the use of data for a more objective and insightful analysis of athletic performance.

Keywords: Machine Learning, Ensemble, IRONMAN 70.3, Athletes Classification, Triathlon

USO DE MACHINE LEARNING ATRAVÉS DE COMITÊS DE ALGORITMOS E MODELOS ISOLADOS PARA CLASSIFICAÇÃO DE ATLETAS DO GÊNERO MASCULINO EM PROVAS DE IRONMAN 70.3 EM NÍVEIS "APENAS CONCLUINTE" E "COMPETITIVO"

1 Introdução

Nos últimos anos, a prática esportiva tem crescido, associada não apenas ao bem-estar físico, mas a busca por socialização e comunidade. A prática tem sido reconhecida por seus inúmeros benefícios, como redução do estresse, perda de peso, melhoria da qualidade de sono, e pelo fator social, tornando-se atrativa para uma nova geração de atletas profissionais e amadores. O relatório *Year in Sport: The Trend Report* (STRAVA, 2024) destaca essa tendência combinada da busca por comunidade, como componente importante na adesão à prática esportiva; do equilíbrio no respeito aos limites do corpo; e do interesse em participar de grandes provas ou eventos.

Um dos esportes que despontam com crescimento é o triatlo, atividade de resistência *(endurance)* que consiste na prática de três exercícios em sequência: natação, ciclismo e corrida. Existem diferentes tipos de triatlo, como o Sprint, Olímpico, IRONMAN 70.3 e IRONMAN. O IRONMAN, marca registrada, é uma competição popular entre os praticantes de triatlo e foi selecionada como alvo deste estudo, especificamente a modalidade "70.3", que consiste em: 1.9 km de natação, 90km de ciclismo e 21.1 km de corrida, em milhas, somando o total de "70.3".

A expansão do interesse em práticas esportivas, bem como a crescente capacidade de processamento e disponibilidade de dados sobre atividades físicas, possibilita a exploração dos dados para geração de *insights*, que permitam aprofundar o conhecimento dos desempenhos na história da prática da atividade levando a comparação e inferência durante a preparação, inclusive servindo de mecanismos para o desenvolvimento de novos meios de empreender através da exploração desses dados, sejam em consultorias atléticas, ou até mesmo no desenvolvimento de aplicativos, onde, de acordo com o relatório da *Fortune Business Insights* (2025), há uma projeção de crescimento do mercado de U$ 93,12 bilhões em 2024 que poderá chegar à U$ 249,05 bilhões em 2032, exibindo um *CAGR (Compound Annual Growth Rate)* de 13,1%.

Dessa forma, a partir de base de dados com registros de provas IRONMAN 70.3 dos anos de 2004 até 2020 (obtida no site *kaggle*), buscou-se investigar como objetivo desta pesquisa a seguinte questão: A aplicação de um comitê de algoritmo melhora a classificação do nível do desempenho de atletas do gênero masculino em provas de uma competição de triatlo quando comparada aos modelos individuais?

Para responder tal questionamento, essa pesquisa procurou, através de uma análise exploratória da base de dados, compreender os possíveis grupos que poderiam ser formados através de estudo dos percentuais de representação, dessa forma construindo as categorias "apenas concluinte" e "competitivo". Após divisão e análise para compreensão da base, foram feitas atividades de pré-processamento, seguidas pelas modelagens dos algoritmos de classificação com treino e teste. Aos resultados individuais foram aplicadas técnicas de comitês de algoritmos, com utilização de métricas clássicas para comparação entre os métodos investigados, a fim de constatar qual obteve o melhor resultado.

2 Fundamentação Teórica

A utilização de algoritmos de aprendizado de máquina na análise de bases de dados estruturadas de resultados de triatlo se apresenta como uma maneira de inovar na compreensão

e classificação de desempenho esportivo, servindo como orientação e exploração de conhecimento em múltiplos níveis, inclusive sendo reflexo de pesquisas científicas vinculadas a análises algorítmicas.

Em um desses estudos, intitulado A Machine Learning Approach to Finding the Fastest Race Course for Professional Athletes Competing in IRONMAN 70.3 Races between 2004 and 2020, Thuany et al. (2023) tiveram como objetivo identificar os percursos de prova mais rápidos para atletas de elite do IRONMAN 70.3, utilizando algoritmos de aprendizagem de máquina (ML).

Nesse estudo, ocorreu a coleta de dados de 16.611 triatletas profissionais de 97 países, competindo em 163 diferentes provas entre 2004 e 2020. Quatro modelos de regressão de ML foram construídos, utilizando gênero, país de origem e localização do evento como variáveis independentes para prever o tempo final da prova. Em todos os modelos, o gênero foi a variável mais importante na previsão dos tempos de conclusão.

O modelo de árvore de decisão simples indicou que os tempos de prova mais rápidos no Campeonato Mundial de IRONMAN 70.3, em torno de 4 horas e 3 minutos, seriam alcançados por homens de países como Áustria, Austrália, Bélgica, Brasil, Suíça, Alemanha, França, Reino Unido, África do Sul, Canadá e Nova Zelândia. Este estudo demonstra a existente oportunidade em explorar técnicas de *machine learning* para analisar grandes volumes de dados de desempenho esportivo e extrair informações, corroborando a relevância da presente pesquisa em explorar a classificação de atletas com base em seus resultados.

Em outro estudo, Thuany et al. (2024) investigaram a influência das diferentes disciplinas (natação, ciclismo e corrida) no tempo total de prova de triatletas profissionais de IRONMAN 70.3. Utilizando a técnica de regressão linear múltipla, o estudo, intitulado Cycling and Running are More Predictive of Overall Race Finish Time than Swimming in Professional IRONMAN 70.3 Triathletes, concluiu que o ciclismo e a corrida são mais preditivos do que a natação.

O estudo analisou 16.611 registros de triatletas que participaram de 787 provas 70.3 entre 2004 e 2020. Utilizando análises de correlação e regressão linear múltipla, os autores buscaram determinar qual disciplina é a mais preditiva para o tempo final da prova. Os resultados indicaram que o ciclismo e a corrida são mais preditivos do tempo total de prova do que a natação, tanto para homens quanto para mulheres. Em homens, os tempos de corrida e natação foram mais amplamente relacionados aos tempos totais de prova do que em mulheres. Além disso, o desempenho na natação mostrou menor disparidade entre os sexos em comparação com o ciclismo e a corrida.

A relevância desse estudo está em considerar as ferramentas tecnológicas atuais, que evoluíram exponencialmente nos últimos anos, para o fornecimento de subsídios à pesquisadores, atletas (profissionais e amadores), entusiastas, e produtores de novos sistemas de produção de tecnologia do esporte que podem usar desse conhecimento para criação ou melhoria de equipamentos relacionados à prática do triatlo.

O *Machine Learning* (ML) é uma ferramenta poderosa para extrair conhecimento e otimizar processos a partir de grandes volumes de dados (McKinsey Global Institute [MGI], 2016). Portanto, sua aplicação é justificada em domínios específicos para gerar novas percepções. No contexto desta pesquisa, o ML é utilizado para analisar o desempenho de atletas masculinos no triatlo IRONMAN 70.3, buscando-se obter inferências valiosas para aprimorar treinamentos, classificar atletas, validar resultados e explorar oportunidades comerciais.

Ao ter realizado a modelagem dos algoritmos de ML, conforme é apresentado no tópico “Método” deste artigo, processando os dados históricos da base utilizada, verificou-se a viabilidade dos modelos, assim como do comitê montado, na classificação de resultados utilizando variáveis independentes (tempos parciais das modalidades e informações etárias)

para prever a variável dependente (categoria de desempenho entre "apenas concluinte" e "competitivo").

Os modelos construídos foram feitos utilizando algoritmos de aprendizado supervisionado de máquina *(supervised machine learning),* sendo-os: *Decision Tree, k-Nearest Neighbors (KNN), XGBoost e Random Forest*.

O k-Nearest Neighbors (KNN) é um método de aprendizagem de máquina supervisionado e não paramétrico que pode ser utilizado tanto para classificação quanto para regressão (Kramer, 2013). Segundo Kramer (2013), o princípio fundamental do algoritmo baseia-se na proximidade, classificando um novo ponto de dados a partir da classe majoritária de seus k vizinhos mais próximos no espaço de características. A simplicidade do KNN em capturar relações complexas em dados o torna uma ferramenta eficaz para o reconhecimento de padrões e a mineração de dados.

A árvore de decisão (Decision Tree) é um modelo versátil empregado em tarefas de regressão e classificação, como a realizada neste estudo. Seu funcionamento consiste em dividir recursivamente os conjuntos de dados em subconjuntos menores com base em regras de decisão simples, formando uma estrutura hierárquica semelhante a uma árvore (Rokach & Maimon, 2008).

Diferentemente das árvores de decisão individuais, a Random Forest (Floresta Aleatória) é um método de *ensemble learning*, ou aprendizagem de conjunto. Conforme explica Breiman (2001), este algoritmo constrói múltiplas árvores de decisão durante o treinamento e agrega seus resultados: para classificação, utiliza a classe mais frequente (moda), e para regressão, a média das previsões. Essa abordagem mitiga o problema de *overfitting* (sobreajuste), tornando o algoritmo eficaz para lidar com grandes conjuntos de dados e alta dimensionalidade (Breiman, 2001).

Por fim, o modelo eXtreme Gradient Boosting (XGBoost) foi utilizado por sua alta performance. Segundo Chen e Guestrin (2016), o XGBoost é um algoritmo de *boosting* que cria novos modelos sequencialmente para corrigir os erros residuais dos modelos anteriores. Seu desempenho é otimizado por meio de técnicas como paralelização, poda de árvores e tratamento eficiente de valores ausentes.

Como forma de avaliar os modelos, foram utilizadas métricas clássicas: *F1-score*, *Precision* (Precisão), *Recall* (Revocação) e *Accuracy* (Acurácia).

O F1-score, como métrica, é calculado como a média de precision e do recall. Esse tipo de medida é passível de utilização em cenários onde ocorre desbalanceamento entre as classes, penalizando modelos que tenham bom desempenho em uma métrica, porém acaba falhando em outra. Ocorre que o F1-score alto implica em significar o modelo como de alta precisão, ou seja, possuindo poucos falsos positivos, assim como alto recall (poucos falsos negativos) (HAND; KIRIELE; CHRISTER, 2023).

Enquanto o precision é responsável por avaliar a identificação de positivas corretas entre todas as identificações positivas feitas pelo modelo (verdadeiros positivos mais falsos positivos), o recall, avalia o número de verdadeiros positivos que foram corretamente identificados pelo modelo em relação a todos os positivos reais (verdadeiros positivos mais falsos negativos). Essas métricas podem possuir alto valor em sistemas de diagnósticos médicos, ou detecção de fraudes (POWERS, 2011).

A acurácia, métrica popularizada e mais intuitiva, na realidade apresenta a previsão de corretos, tanto em relação à verdadeiros positivos quanto verdadeiros negativos, no conjunto do número total de previsões realizadas. De outra forma, a acurácia medirá se o modelo foi exato em uma percepção geral. Por mais simples que seja de entender a acurácia, deve-se ter atenção aos cenários com classes desbalanceadas, pois um modelo pode apresentar acurácia em altos níveis por simplesmente prever classes majoritárias (POWER, 2011).

Dessa forma, considerando os conceitos dos modelos apresentados de aprendizagem de máquina não supervisionado, e que o objeto dessa pesquisa se preocupou com uma classificação de performance (“apenas concluinte” e “competitivo”), com base em variáveis independentes, esses foram utilizados, assim como também postos em modelos *ensemble*, e para avaliar o desempenho desses algoritmos, foram utilizadas as métricas aqui mencionadas para análises e comparações.

3 Metodologia

A pesquisa desse artigo se apresenta como de caráter exploratória, visto que buscou-se investigar, analisar e comparar diferentes métodos de classificação (individuais e comitê algoritmos) para uma atividade que tem crescido mundialmente, a prática de triatlo da modalidade IRONMAN 70.3, através de resultados históricos do evento durante um período de 16 anos.

Assim como, possui uma abordagem empírica, visto que o estudo foi feito com base em dados reais, além dos algoritmos em seus modelos individuais e em comitê, onde foram testados e avaliados por meio de experimentação direta com os dados coletados, passando por processamento e análise quantitativa dos resultados.

4 Utilização de Inteligências Artificiais Generativas (IAGs) como ferramentas de suporte à modelagem

Com o advento dos variados modelos de Inteligência Artificial Generativa, tem sido possível para àqueles que não possuem o domínio técnico de alguma linguagem de programação, ou que possua conhecimentos básicos, utilizar dessa ferramenta como um meio para atingir um determinado fim.

Dessa forma, foram utilizadas as ferramentas GPT-4.1 e Gemini como meios de codificar linhas mais complexas quando os autores não conseguiam produzi-las, ou, ao produzi-las, acabavam com erros repetidos, servindo essas IAGs (Inteligências Artificiais Generativas) como agentes de auxílio.

Essas ferramentas também foram importantes ao “plotar” gráficos, tabelas e outras informações durante o momento de análise exploratória.

Os códigos resultados desse trabalho estarão disponíveis no GitHub.

Portanto, através desses mecanismos de IAG, foi possível chegar a construção dos modelos de algoritmos de forma isolada e em estrutura de comitê.

5 Coleta e pré-processamento

Os dados utilizados nesta pesquisa foram obtidos de uma base de dados pública disponível no site kaggle (link disponível nas referências), contendo registros de provas de IRONMAN 70.3 realizadas entre os anos de 2004 e 2020. A base de dados inclui informações sobre o desempenho dos atletas, assim como outras colunas descritas abaixo:

Tabela 1. Metadados da base de dados

Coluna	Atribuição
Gender	M: Masculine (Masculino) F: Feminine (Feminino)

AgeGroup	"Grupo etário" - Ex.: Grupo entre "40-44"
AgeBand	"Faixa etária" - Ex.: 40
Country	"País" - Ex.: Andorra
CountryISO2	Código do país - Ex.: Andorra - AD
EventYear	"Ano do evento"
EventLocation	"Local do evento"
SwimTime	"Tempo de natação"
Transition1Time	"Tempo da transição 1"
BikeTime	"Tempo de ciclismo"
Transition2Time	"Tempo da transição 2"
RunTime	"Tempo de corrida"
FinishTime	"Tempo de conclusão da prova"

Fonte: Produzida pelos próprios autores

Os tempos parciais das provas de natação, ciclismo, corrida, transição (T1 e T2), além do tempo de conclusão da prova já foram extraídos da kaggle de forma convertida em segundos, considerando que foi utilizada a linguagem *python* para codificação e o *Google Colab* como ambiente de processamento.

Após acesso a base, o pré-processamento seguiu o fluxo, conforme a figura 1, para análise prévia da base:

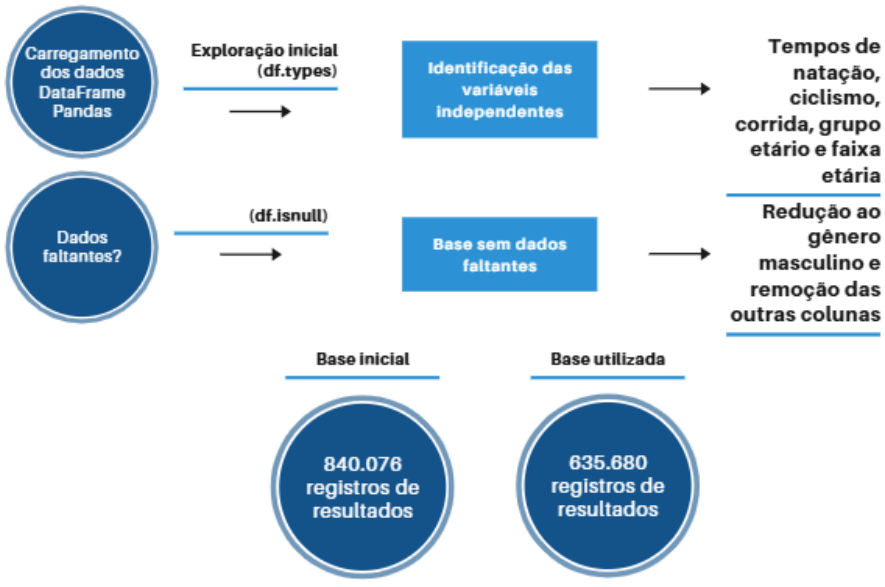

Figura 1
Fluxograma de pré-processamento da base
Nota: O fluxograma detalha as etapas de tratamento dos dados, desde o carregamento e exploração inicial até a filtragem que resultou na base final utilizada para análise. Figura elaborada pelos autores na ferramenta Canva.

Ocorre que, ao verificar quais eram os grupos e faixas etárias existentes na base, verificou-se a existência de valores zero, conforme apontado em imagem abaixo:

```
Categorias disponíveis em AgeGroup:
['40-44' '45-49' '35-39' '50-54' '25-29' '18-24' '30-34' '55-59' '00'
 '60-64' '70-74' '65-69' '75-79' '80-84' '85-89']

Valores únicos em AgeBand: [40 45 35 50 25 18 30 55  0 60 70 65 75 80 85]

Distribuição de Performance:
Performance
Apenas Concluinte    508497
Competitivo          127183
Name: count, dtype: int64
```

Figura 2
Valores das colunas de grupo etário e faixa etária
Nota: A figura apresenta a distribuição dos valores de grupo etário e faixa etária. Figura elabora pelos próprios autores retirando da ferramenta Google Colab.

Considerando que os valores de idade do grupo e faixa etária são considerados nessa pesquisa como variáveis independentes, decidiu-se pela remoção dos resultados que apresentassem os valores “00”, e “0”, conforme a figura 2 acima, visto que poderiam resultar em distorções na compreensão dos modelos *(outliers)*.

Além das atividades acima ilustradas para manutenção da qualidade de dados, foi necessária uma análise para criação das classes de desempenho mediante os resultados existentes na base, para tanto, foi adotado o critério de porcentagens. Os 20% dos atletas com os melhores tempos finais de prova foram classificados como “Competitivos”, enquanto os demais foram enquadrados como “Apenas Concluinte”.

Tabela 2. Resultado após aplicação das porcentagens

Classificação	Quantidade de registros
Competitivos	117.118
Apenas Concluinte	508.280

Fonte: Produzida pelos próprios autores

A segmentação do desempenho com base em porcentagens é uma abordagem metodológica reconhecida que permite uma classificação relativa e adaptável a diferentes amostras de atletas (Kraemer & Blasey, 2015; Schorer et al., 2017). Seguindo essa abordagem, os participantes foram classificados em dois grupos: competitivo (20%, n = 117.118) e apenas concluinte (80%, n = 508.280), cuja distribuição está apresentada na Tabela 2.

Ademais, para validar a utilização dessa distribuição de porcentagem, foi realizada a visualização dos tempos registrados no tempo total da prova (FinishTime), segmentado tanto por grupo etário (AgeGroup) quanto pela classificação da performance (competitivo e apenas concluinte) através de gráfico de *boxplot*, abaixo:

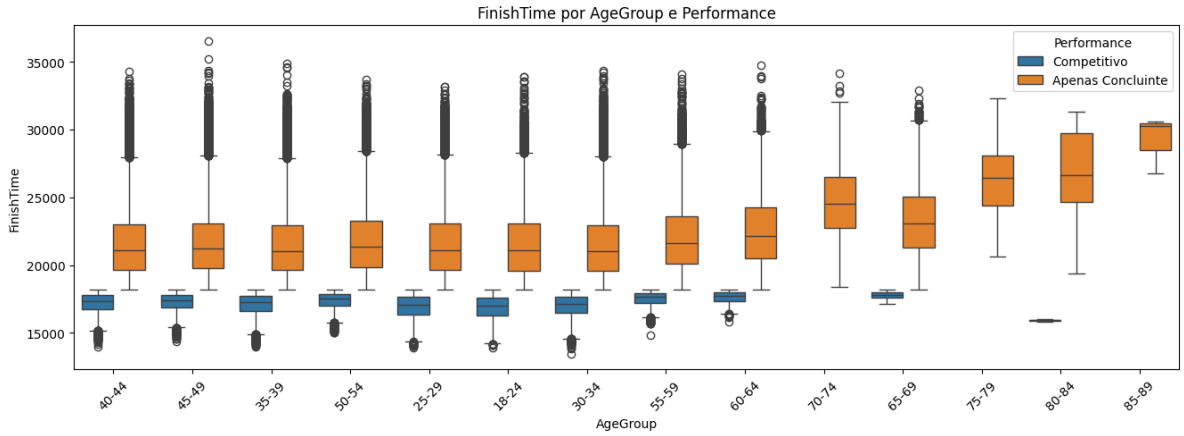

Figura 3.
Boxplot dos Tempos
Nota: Bloxpot das variações de classificação. Produção pelos próprios autores por codificação em python no Google Colab.

É notável, através do gráfico da figura 3, que para todas as idades, os atletas que foram classificados como competitivo apresentam tempo de conclusão menores em comparação aos apenas concluintes, o que é um indício que valida o critério de classificação adotado. Ademais, observa-se que a mediana dos tempos competitivos se mantém mais baixa e estável nas faixas de 18 até 44 anos, sendo possível inferir que o ápice do desempenho esportivo se concentra nesse intervalo etário. Da mesma forma, também se infere que esse desempenho reduz a partir das faixas etárias de 45 anos, pois há um aumento gradativo nos tempos medianos, o que pode evidenciar o impacto fisiológico da idade sobre o desempenho.

É importante destacar, em relação ao gráfico, uma maior variabilidade nos tempos do grupo apenas concluinte, o que pode indicar diferentes níveis de preparação, condicionamento físico e/ou experiência. Além do mais, a partir de faixas etárias mais avançadas, como acima dos 65 anos de idade, é possível perceber a redução da quantidade de participantes mediante a redução das medianas.

Ocorre que, ao analisar o gráfico de boxplot, e a tabela de resultados por grupo etário, detectou-se outliers, onde apenas dois competidores foram classificados em nível competitivo no grupo etário de 85-89 anos. Decidiu-se pela remoção desses resultados da base, pois os pesquisadores entenderam que podiam interferir no treinamento dos modelos.

6 Resultados

Conforme apontado acima na figura 1, utilizou-se como variáveis independentes os tempos de natação (SwimTime), ciclismo (BikeTime) e corrida (RunTime), assim como grupo etário e faixa etária, não sendo considerado o tempo de conclusão.

Um dos primeiros aspectos importantes a ser mencionado em relação à resultados desse estudo está na desconsideração de utilização da variável independente FinishTime. Tal fato se deu devido ao primeiro momento em que se utilizou o algoritmo árvore de decisão, onde se teve como resultado um único nó de decisão que foi baseado no tempo de conclusão, mesmo com as outras variáveis sendo utilizadas. Portanto, houve uma segmentação determinística, na separação entre atletas em nível competitivo ou apenas concluinte.

Contudo, no intuito de melhor explorar o potencial dos modelos, ignorou-se a variável FinishTime, focando nos tempos de conclusão das disciplinas que compõem o triatlo IRONMAN 70.3.

Ademais, objetivando explorar melhor os comitês ensemble, em que ocorre a combinação das predições individuais por meio de mecanismo de votação, utilizou-se duas abordagens comuns para ações de classificação, sendo-as *hard voting* e o *soft voting* (Gerón, 2019).

No sistema de votação hard voting, ou de votação majoritária *(majority voting)*, encontra-se um método de agregação simples e direto, em que a predição final considerada é uma nova instância de dados da classe que recebe o maior número de votos dos modelos individuais dos quais compõem o comitê (Rokach, 2010).

Por sua vez, o soft voting conta com uma abordagem que infere mais certeza, visto que leva em consideração a confiança de cada modelo em sua predição, ou seja, em vez de votar em uma classe, cada classificador atribui uma probabilidade de pertencimento a cada uma das classes possíveis, seguindo pela agregação das probabilidades (cálculo da média) para cada classe e seleciona como predição final a classe que obtiver a maior probabilidade média (Gerón, 2019).

De forma visual, apresenta-se a sequência de tabelas os resultados, inicialmente, com a acurácia de cada modelo e comitês na tabela 3, abaixo:

Tabela 3. Índice de acurácia dos modelos individualizados e comitês

Modelo	Acurácia
Decision Tree	0.98
KNN	0.98
XGBoost	0.98
Random Forest	0.98
Comitê com Hard Voting	0.98
Comitê com Soft Voting	0.98

Fonte: Produzida pelos próprios autores

Resultados em relação à classificação “Apenas Concluinte” na tabela 4:

Tabela 4. Índice de métricas dos modelos individualizados e comitês para “apenas concluinte”

Modelo	Precision	Recall	F1-score
Decision Tree	0.99	0.99	0.99
KNN	0.99	0.99	0.99
XGBoost	0.99	0.99	0.99
Random Forest	0.99	0.99	0.99
Comitê com Hard Voting	0.99	0.99	0.99
Comitê com Soft Voting	0.99	0.99	0.99

Fonte: Produzida pelos próprios autores

Para “competitivo” na tabela 5:

Tabela 5. Índice de métricas dos modelos individualizados e comitês para “competitivo”

Modelo	Precision	Recall	F1-score
Decision Tree	0.96	0.96	0.96
KNN	0.97	0.97	0.97
XGBoost	0.99	0.99	0.99
Random Forest	0.97	0.97	0.97
Comitê com Hard Voting	0.97	0.96	0.97
Comitê com Soft Voting	0.97	0.97	0.97

Fonte: Produzida pelos próprios autores

7 Discussão

A análise dos resultados obtidos pelos modelos individuais e pelos comitês de votação (hard voting e soft voting) evidencia bons resultados da capacidade de classificação em todos os casos. Os valores de acurácia, precisão, recall e f1-score permaneceram consistentemente acima de 95% para ambas as classes (competitivo e apenas concluinte), indicando que os modelos de algoritmos selecionados conseguem trabalhar de forma satisfatória quando avaliados de forma isolada.

É possível inferir, através desses resultados que, as variáveis independentes adotadas (tempos de conclusão de natação, ciclismo, corrida, grupo etário e faixa etária), mediante a distribuição em percentis, fez com que se tivesse uma discriminação precisa para o problema de classificação proposto.

Ao comparar o desempenho dos comitês com o dos modelos individuais, observou-se praticamente inexistente. Embora os comitês, tanto por hard quanto por soft voting, tenham apresentado uma leve redução nos erros (falsos positivos e falsos negativos) em comparação com os modelos individuais, é uma melhoria de baixo nível.

Compreende-se que isso possa ocorrer pelo fato de que os modelos individuais já demonstraram desempenhos altos, indicando que o problema em questão apresenta uma separabilidade suficiente para a maioria dos algoritmos tradicionais de classificação, mesmo sem a necessidade de técnicas de ensemble mais complexas.

A adoção de um comitê de modelos (ensemble) é uma estratégia validada para mitigar os erros individuais de cada algoritmo, aumentando assim a robustez da predição final. Em aplicações práticas, como em tecnologias para o planejamento criterioso do treinamento de atletas, o uso de um ensemble adiciona uma camada de credibilidade e segurança, pois a decisão agregada é inerentemente mais confiável do que a de qualquer modelo isolado.

A abordagem de soft voting, em particular, se destaca por incorporar o grau de confiança das predições dos modelos, o que pode ser útil em situações-limite ou em cenários com maior ambiguidade entre as classes.

Contudo, considerando os resultados observados, a escolha entre um modelo individual ou o comitê pode ser pautada também pela simplicidade operacional, desempenho computacional e facilidade de manutenção do *pipeline*. Para este problema específico, a boa performance dos modelos individuais sugere que soluções mais simples podem ser igualmente

eficazes, dependendo dos requisitos de implementação e recursos disponíveis, mediante a proposta de classificação utilizada neste estudo.

8 Limitações e pesquisas futuras

Apesar da performance alcançada pelos modelos e comitês testados, devem ser apresentadas limitações em relação a esse estudo, que podem ser oportunidades de melhorias e novas pesquisas passíveis de exploração.

O primeiro ponto está no critério de classificação, baseado unicamente em cortes de porcentagem do tempo total, o que pode não refletir nuances técnicas ou contextuais das provas. O não emprego de variáveis demográficas ou ambientais restringe a generalização do modelo, especialmente para outras populações ou eventos.

O estudo também teve uma análise limitada dos erros, levando em conta a percepção de que em provas de resistência, como o IRONMAN 70.3, é comum a presença de tempos discrepantes, contudo é uma análise que não considera outras possibilidades, o que pode gerar possíveis problemas de *fairness* (equidade) ou subgrupos menos representados podem passar despercebidos.

Por fim, a ausência de análise detalhada dos erros, assim como a simplicidade do método de comitê adotado, indica caminhos para pesquisas futuras visando robustez e aplicabilidade ampliada, inclusive havendo a possibilidade de explorar outros cenários de previsão, que não esteja só focado apenas no pós-prova.

9 Conclusão

Esta pesquisa explorou a aplicação de modelos de aprendizado de máquina, tanto individualmente quanto em comitês (ensemble), para classificar atletas de triatlo IRONMAN 70.3 em níveis de conclusão competitivo e apenas concluinte. Os resultados demonstraram que todos os modelos testados, apresentaram um desempenho de classificação alto, com métricas de acurácia, precisão, recall e F1-score consistentemente acima de 95%.

A comparação entre os modelos individuais e os comitês de hard voting e soft voting revelou que, embora os comitês tenham proporcionado uma leve melhoria na redução de erros, a diferença de desempenho foi, relativamente, baixa. Isso indica que as variáveis de tempo das modalidades (natação, ciclismo e corrida), e as relacionadas às idades dos competidores do gênero masculino, possuem um poder preditivo muito forte para o problema de classificação proposto, tornando a separabilidade entre as classes bastante clara para os algoritmos.

Em suma, a aplicação de comitês de algoritmos, embora não tenha gerado um ganho substancial de desempenho em relação aos modelos individuais neste contexto específico devido à performance inicial, ainda representa uma estratégia válida para aumentar a robustez e mitigar erros pontuais em cenários de classificação de desempenho esportivo. A escolha entre um modelo individual e um comitê dependerá de fatores como a complexidade do problema, os requisitos de robustez da decisão e a facilidade de implementação e manutenção.

Referências

Analysis of Ironman 70.3 races from 2004 to 2020. (2022). Kaggle. https://www.kaggle.com/code/aiaiaidavid/analysis-of-ironman-70-3-races-from-2004-to-2020

BREIMAN, L. (2001). **Random Forests. Machine Learning**, 45(1), 5-32, https://doi.org/10.1023/A:1010933404324.

Chen, T., & Guestrin, C. (2016). **XGBoost: A scalable tree boosting system.** In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Fortune Business Insights. (2024). **mHealth market size, share & industry analysis, by category, by service type, by service provider, and regional forecast, 2024-2032**. https://www.fortunebusinessinsights.com/pt/industry-reports/mhealth-market-100266

Gerón, A. (2019). **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems** (2nd ed.). O'Reilly Media. 182-184.

Hand, D. J., Kiriele, N., & Christer, P. (2023). **A review of the F-measure: Its history, properties, criticism, and alternatives**. ACM Computing Surveys, 56(3), 1–37. https://doi.org/10.1145/3606367

Kraemer, W. J., & Blasey, M. (2015)**. How many subjects? Statistical power analysis in research**. Human Kinetics.

Kramer, O. (2013). **K-nearest neighbors**. In Dimensionality reduction with unsupervised nearest neighbors (pp. 13–23). Springer. https://doi.org/10.1007/978-3-642-38652-7_2

McKinsey Global Institute. (2016). **The age of analytics: Competing in a data-driven world.** McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20analytics/our%20insights/the%20age%20of%20analytics%20competing%20in%20a%20data%20driven%20world/mgi-the-age-of-analytics-full-report.pdf

Powers, D. M. W. (2011). **Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation.** Journal of Machine Learning Technologies, 2(1), 37–63.

Rokach, L. (2010). **Ensemble-based classifiers.** Artificial Intelligence Review, 33(1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7

Rokach, L., & Maimon, O. (2008). **Decision trees.** In Data mining and knowledge discovery handbook (pp. 165–192). Springer. https://doi.org/10.1007/0-387-25465-X_9

Schorer, J., Büsch, D., & Fischer, L. (2017). **Talent identification and development in sport: Current state and future directions.** Journal of Sports Sciences, 35(10), 969–976. https://doi.org/10.1080/02640414.2016.1198226

Strava. (s.d.). Strava releases annual year in sport trend. https://press.strava.com/pb/articles/strava-releases-annual-year-in-sport-trend

Thuany, M., Costa, T. H., Arriel, R. A., Souza-Silva, W., & Costa, V. P. (2023**). A machine learning approach to finding the fastest race course for professional athletes competing in Ironman® 70.3 races between 2004 and 2020.** International Journal of Environmental Research and Public Health, 20(4), 3619. https://doi.org/10.3390/ijerph20043619

Thuany, M., Costa, T. H., Arriel, R. A., & Costa, V. P. (2024). **Cycling and running are more predictive of overall race finish time than swimming in professional Ironman® 70.3 triathletes.** Frontiers in Sports and Active Living, 6, 1214929. https://doi.org/10.3389/fspor.2024.1214929