UNIVERSIDADE da MADEIRA
CDA
cidesd
INTERACTIVE TECHNOLOGIES INSTITUTE
ITI-LX UIDP/50009/2020 - IST-ID
fct Fundação para a Ciência e a Tecnologia

INTERNATIONAL SEMINAR

SPORT AND SCIENCE 2025

10-12 de ABRIL

LIVRO DE ATAS

FUNCHAL
Região Autónoma da Madeira
Secretaria Regional de Educação, Ciência e Tecnologia
Direção Regional de Desporto

Seminário Internacional Desporto e Ciência 2025

Comissão Organizadora
Rui Trindade, Helder Lopes, Élvio Gouveia, Catarina Fernando, Ana Luísa Correia, Ricardo Alves, Hélio Antunes, Cíntia França, João G. Saldanha, Francisco Teixeira, João Martins, Francisco Santos.

Comissão Científica
Élvio Gouveia, Helder Lopes, Rui Ornelas, Duarte Freitas, Catarina Fernando, Jorge Soares, Ana Rodrigues, Ricardo Alves, Hélio Antunes.

Ficha Técnica

Livro de Atas Seminário Internacional Desporto e Ciência 2025
ISBN: 978-989-9230-11-8

Coordenação da Edição: Hélio Antunes, Rui Ornelas, Helder Lopes, Catarina Fernando, Ana Rodrigues, Ana Luísa Correia, Cíntia França & Élvio Gouveia.

Editor: Universidade da Madeira
2025 – Funchal, Portugal
Suporte: Eletrónico
Formato: PDF/PDF/A

Athlete Security Management: the case of the Madeira Island Ultra Trail Event

Eduardo Marques[1,2], Ricardo Pinheiro[1]

(1) University of Madeira, (2) CiTUR Madeira
emarques@uma.pt, 2042416@student.uma.pt

Introduction

Trail running is defined as a running sport conducted in outdoor, mountainous terrain, often involving steep ascents and descents, with some races exceeding 24 hours in duration (Scheer et al., 2020). The sport has grown significantly in popularity, with around 9.1 million participants in the U.S. in 2017. By 2021, the International Trail Running Association (ITRA) had registered 1.8 million runners from 163 countries using its Performance Index (PI) and listed nearly 8,000 trail races in its event calendar (Crawley, 2021), fostering sport and also tourism (Perić and N. Slavić, 2019). Predicting a runner's checkpoint arrival time is crucial for race organizers, as delays may indicate potential danger, enabling timely emergency response measures.

Study Objective

The objective of this work was to develop a more accurate method for predicting athletes' arrival times at each checkpoint. Traditional velocity-based algorithms often rely on simplistic assumptions, such as constant speed or linear progression, which fail to account for dynamic factors like fatigue, elevation changes, and weather conditions. To address these limitations, we leveraged Machine Learning (ML) techniques, based on the work of (Fogliato et all, 2021), analysing historical race data and individual performance patterns to generate better adaptive predictions. By improving checkpoint time estimates, this approach enhances race safety, resource allocation, and logistical planning for organizers.

Methodology

To achieve the objective of accurately predicting athletes' arrival times at checkpoints during trail races, we proposed a data-driven methodology, combining traditional race analytics with supervised machine learning (ML) techniques, including historical race and athlete's data, engeneering features based on distance and elevation, with some normalization of data, removing erroneous ou missing athlete data. The use of Random Forest (Breiman, 2001) and Lasso (Tibshirani, 2009) Machine Learning models will be compared with simple velocity models. The training will be incremental, using 6 years of historical data and continuos analysis of results. The evaluation metric used will be Mean Absolute Error (MAE) (Hastie et all, 2009).

Results

Initial testing of simple velocity-based models revealed that average speed across the entire race provided more accurate predictions than using only the most recent checkpoint velocities: Average speed model with 1,501 seconds MAE (mean absolute error), Last checkpoint velocity with 1,807 seconds MAE, and, Last two checkpoints' average velocity with 1,724 seconds MAE. This

confirms that long-term pacing trends are more reliable than short-term fluctuations in predicting arrival times.
The initial ML model, trained solely on distance-based features (e.g., segment length, elevation), underperformed compared to the baseline average-speed model. However, after incorporating athlete-specific features, the model's accuracy improved significantly. The Random Forest (RF) Model achieved an MAE of 1,134 seconds, a 24% improvement over the best baseline model.

Conclusions
The application of the Random Forest machine learning model significantly enhanced race performance predictions, achieving an average improvement of 24% (6.1 minutes) compared to baseline methods. This approach leveraged historical race data and athlete performance metrics, enabling more accurate time estimations by identifying key patterns and relationships within the dataset. However, the study encountered limitations related to athlete data consistency. Specifically, tracking the same athlete across multiple races proved challenging due to inconsistencies in data recording, missing entries, or variations in athlete identification. This limitation may have affected the model's ability to fully capture individual performance trends over time.
As future improvements we propose a more robust athlete tracking systems to ensure consistent data across competitions, and expand the dataset with additional features such as weather conditions, course difficulty, and real-time physiological data (e.g., heart rate, GPS tracking) to further refine predictions. Also, another line of research could explore ML models such as, XGBoost and Neural Networks to assess whether different algorithms yield even better accuracy.
Despite these challenges, the results demonstrate that machine learning, particularly Random Forest, holds strong potential for performance prediction in endurance sports, provided that data quality and athlete tracking are further optimized.

Keywords: Machine Learning, Random Forest, Trail Running, Prediction

References:
V. Scheer, P. Basset, N. Giovanelli, G. Vernillo, G. P. Millet, and R. J. S. Costa, "Defining Off-road Running: A Position Statement from the Ultra Sports Science Foundation," Int J Sports Med, vol. 41, no. 5, pp. 275–284, May 2020, doi: 10.1055/a-1096-0980.

Bob Crawley, Trail running and ITRA report, August 2021, Available: https://itra.run/content/news/EN-TRAIL_RUNNING_REPORT_2021.pdf

R. Fogliato, N. L. Oliveira, and R. Yurko, "TRAP: a predictive framework for the Assessment of Performance in Trail Running," Journal of Quantitative Analysis in Sports, vol. 17, no. 2, pp. 129–143, 2021.

M. Perić and N. Slavić, "Event sport tourism business models: The case of trail running," Sport, Business and Management: An International Journal, vol. 9, no. 2, pp. 164–184, 2019.

L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

T. Hastie, R. Tibshirani, and J. Friedman, “Overview of Supervised Learning,” in The Elements of Statistical Learning: Data Mining, Inference, and Prediction, T. Hastie, R. Tibshirani, and J. Friedman, Eds., in Springer Series in Statistics. , New York, NY: Springer, 2009, pp. 9–41. doi: 10.1007/978-0-387-84858-7_2.

R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.