Weather-Aware Prediction of Trail Running Finish Times Using Machine Learning

Muhammad Annas Darunnaja[1], Mokhamad Amin Hariyadi[2], Okta Qomaruddin Aziz[3], Johan Ericka Wahyu Prakasa [4], Agung Teguh Wibowo Almais[5]
[1,2,3,4,5]Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia

Article Info

Article history:
Received May 05, 2026
Revised May 27, 2026
Accepted June 06, 2026

Keywords:
Trail running
Extreme Gradient Boosting
Explainable AI
Event-based modeling
Weather Impact

ABSTRACT

This study investigates the role of environmental variables in improving the prediction of Mean Finish Time (MFT) in trail running events. While previous approaches primarily rely on track-related features to predict individual athlete performance, the contribution of dynamic weather conditions at the event level remains insufficiently explored. This research adopts a quantitative modeling approach using Extreme Gradient Boosting (XGBoost) to analyze 36,700 race records integrated with spatio-temporal weather data. To rigorously prevent data leakage, a controlled experimental design was implemented using Group Shuffle Split based on race titles, comparing a model that incorporates environmental variables against one relying solely on track characteristics. The results show that the inclusion of weather variables significantly improves predictive reliability, reducing the Mean Absolute Percentage Error (MAPE) from 10.05% to 8.50% and increasing the coefficient of determination (R^2) to 0.7836. Further analysis reveals that environmental variables, particularly temperature, interact with terrain difficulty and disproportionately influence high-effort events. In conclusion, integrating environmental variables significantly enhances predictive accuracy, offering a novel, data-driven approach for race organizers to calculate ideal Cut-Off Times (COT) and Cut-Off Points (COP) based on dynamic environmental constraints.

Corresponding Author:
Muhammad Annas Darunnaja,
Department of Informatics Engineering,
Jl. Gajayana No. 50, Kec. Lowokwaru, Malang City, East Java 65144.
Email: 220605110130@student.uin-malang.ac.id

1. INTRODUCTION

Trail running is an endurance sport conducted on natural terrains characterized by heterogeneous surfaces, elevation variability, and dynamic environmental conditions [1]. Unlike road running, race duration in trail running is not solely determined by distance but is strongly influenced by complex interactions between terrain characteristics and environmental factors such as temperature, precipitation, and wind [2], [3]. As global participation in trail running continues to grow, accurate estimation of race duration has become increasingly essential for both athletes and event organizers. Mean Finish Time (MFT) serves as a practical macro-level indicator of event difficulty, supporting decisions related to pacing strategies, cut-off times, and logistical planning [3].

Recent studies have applied machine learning to predict endurance running performance, with ensemble methods demonstrating consistent superiority over traditional statistical models in handling structured tabular data [4], [5], [6]. However, existing work has primarily focused on individual athlete performance in ultramarathons [2], [7], leaving the event-level prediction problem and in particular the contribution of dynamic environmental conditions insufficiently explored. Weather variables such as temperature, precipitation, and wind speed are known to influence endurance performance [8], [9], [10], yet their systematic integration into event-level predictive models remains limited.

This study addresses that gap by shifting the analytical paradigm from individual-level to event-level prediction, with the specific objective of isolating and quantifying the contribution of environmental variables to MFT prediction accuracy. Rather than benchmarking multiple algorithms, a single well-established model [11], is deliberately employed so that any observed performance differences can be attributed strictly to the inclusion of weather features. In addition, Explainable Artificial Intelligence (XAI) techniques specifically SHAP (SHapley Additive exPlanations) [12], are applied to provide interpretable insights into how environmental variables modulate predicted race durations under varying thermal and meteorological loads [13].

The primary contribution of this study is the shift from individual-level prediction toward an event-level analytical framework from individual athlete performance prediction toward a comprehensive event-level predictive framework that systematically evaluates environmental impacts on trail running finish times. While prior studies predominantly focus on athlete-specific physiological or historical performance indicators, this research integrates dynamic weather variables with structural terrain characteristics to quantify their combined influence on Mean Finish Time (MFT) prediction reliability.

Furthermore, this study emphasizes methodological rigor by implementing a leakage-controlled evaluation strategy [14], using Group Shuffle Split based on race titles, ensuring that recurring race events across different years do not simultaneously appear in training and testing partitions. This approach enables a more realistic assessment of model generalization under unseen event conditions. The primary scientific contributions of this study are summarized as follows:

1. Weather-Aware Event-Level Prediction Framework: Proposing a machine learning framework that integrates spatio-temporal environmental variables with structural trail difficulty metrics to predict Mean Finish Time at the event level.
2. Leakage-Controlled Validation Strategy: Implementing a strict Group Shuffle Split validation mechanism [14] based on unique race identities to prevent event memorization and improve experimental reliability.
3. Quantification of Environmental Contributions: Estimating the relative contribution of weather variables through Explainable Artificial Intelligence (TreeSHAP) [12], [13], enabling interpretable analysis of how environmental conditions modify predicted race durations.
4. Practical Decision-Support Implications: Providing a data-driven framework that can support race organizers in dynamically adjusting Cut-Off Times (COT) and Cut-Off Points (COP) under varying environmental conditions.

2. RESEARCH METHOD

This study adopts a structured machine learning pipeline consisting of data acquisition, preprocessing, feature engineering, model training, and evaluation[14]. The primary objective is not merely to develop a predictive model, but to systematically evaluate the contribution of environmental variables to Mean Finish Time (MFT) prediction. To achieve this, multiple experimental scenarios were designed to isolate the impact of weather features on model performance. The implementation of the proposed model is available in a public repository on github for reproducibility purposes [15].

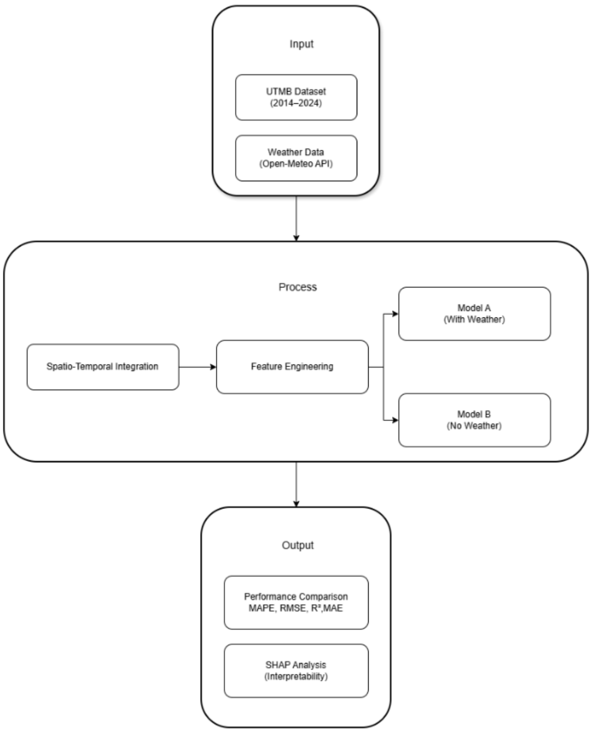

Figure 1. Weather-Aware Framework for Trail Running Finish Time Prediction

2.1. Data Acquisition and Preprocessing

The dataset used in this study was obtained from the UTMB World Race Data Sheet (2014–2024), comprising 38,398 event-level records [16]. To incorporate environmental influence, historical weather data including maximum temperature, minimum temperature, precipitation, and wind speed were retrieved using the Open-Meteo API based on event location and date [17]. After data cleaning and validation, 36,700 records were retained for analysis. For analytical clarity, the dataset variables were grouped into four primary categories:

1. Structural Terrain Features: (*distance_km, elevation_gain_m, km_effort, eg_per_km, elevation_variance*).
2. Environmental Weather Features: (*wx_temp_max, wx_temp_min, wx_temp_mean, wx_precip_sum, wx_rain_sum, wx_wind_speed*).
3. Participation and Demographic Features: (*n_participants, n_women, n_countries*).
4. Geographical and Spatial Features: (*location_elevation_m, longitude, latitude, continent, country*).

Preprocessing was conducted to ensure model reliability and prevent data leakage. The following steps were applied:

1. Removal of post-race variables (e.g., winning time, DNF counts) to avoid leakage [14].
2. Filtering invalid values (e.g., negative distances or elevation).
3. Handling missing values using XGBoost's inherent sparsity-aware mechanism [11].
4. Encoding categorical features where necessary.

All preprocessing steps were implemented using the scikit-learn framework [18].

2.2. Feature Engineering

To represent track structural difficulty, the *km_effort* feature was constructed based on ITRA guidelines [19]. Conceptually, *km_effort* standardizes the combined horizontal distance and vertical elevation gain into a single comparable metric, allowing consistent difficulty assessment across diverse race profiles. The feature is calculated as follows, where *EG* denotes total elevation gain in meters:

$$km\_effort = distance + \frac{EG}{100}$$

Environmental variables were incorporated as primary weather predictors, comprising *wx_temp_max*, *wx_temp_min*, *wx_precipitation*, and *wx_wind_speed*. A complete description of all features used in both models, including their data types and sources, is provided in Table 3.
To stabilize variance and reduce the right-skewed distribution of the target variable, MFT was log-transformed prior to model training [20]:

$$y^* = \log(1 + y)$$

This transformation converts the skewed MFT distribution into a near-Gaussian form, improving regression stability while preserving extreme values commonly observed in ultra-endurance events.

Model performance was evaluated using four complementary metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2) [21], and Mean Absolute Percentage Error (MAPE) [22]. MAE and MAPE provide intuitive, scale-aware measures of average prediction deviation, while RMSE penalizes large errors more heavily making it particularly sensitive to the extreme outliers present in ultra-endurance event durations. R^2 quantifies the proportion of variance explained by the model, offering a global measure of predictive reliability. Together, these metrics provide a comprehensive and balanced assessment of model accuracy and robustness.

2.3. Experimental Design

To comprehensively analyze feature contributions, two experimental models were developed: Model A (full feature set, including weather variables) and Model B (track characteristics only, excluding weather variables). To rigorously evaluate both models and mitigate the risk of event-level data leakage, the dataset was partitioned using a Group Shuffle Split strategy based on race_title (90:10 ratio) [18]. Here, race_title is operationally defined such that identical event names across different years are treated as a single, unique group, ensuring that historical records from the same race event do not appear simultaneously in both the training and test sets. This grouping prevents the model from memorizing recurring race routes rather than learning generalizable environmental patterns.

XGBoost [11] was selected as the modeling framework for both models. Compared to alternative gradient boosting implementations such as LightGBM and CatBoost [5], [23], XGBoost offers robust handling of tabular data sparsity, competitive execution speed, and well-documented native integration with TreeSHAP [12], which is central for the interpretability analysis. As the primary objective of this study is to evaluate the contribution of weather features rather than to benchmark algorithms, a single, well-established model was deliberately employed to ensure that any observed performance differences are strictly attributable to changes in feature composition.

Hyperparameter tuning was performed exclusively on the training set using GroupKFold cross-validation (5 splits) integrated with RandomizedSearchCV [14], [18], with race_title as the grouping key to maintain consistency with the partitioning strategy described above.

3. RESULTS AND DISCUSSION

3.1. Hyperparameter Tuning and Preprocessing Results

The log-transformation successfully converted the right-skewed target distribution into a near-Gaussian distribution, significantly improving the stability of the regression model. The hyperparameter optimization process for Model A, evaluated via 5-fold cross-validation, identified the following optimal configuration: learning_rate: 0.05, max_depth: 8, n_estimators: 1000, min_child_weight: 3, subsample: 0.8, and colsample_bytree: 0.8. The optimal hyperparameters identified for Model A were uniformly applied to Model B to ensure that any performance variance was strictly attributable to changes in feature composition rather than algorithmic tuning.

3.2. Comparative Analysis of Weather vs Non-Weather Models

To rigorously address the risk of event-level data leakage, the models were evaluated using a strict Group Shuffle Split based on race titles. This approach prevents the model from memorizing recurring race routes across different years. Through empirical testing of multiple data partitioning configurations (e.g., 80:20 and 70:30), a 90:10 split ratio was deliberately selected as the optimal configuration. While all split ratios demonstrated a consistent predictive advantage when incorporating weather data, The 90:10 split was selected to maximize training diversity while preserving an independent evaluation set.

Furthermore, it is important to contextualize the choice of evaluation metrics. Previous predictive models focusing on specific, single-category race distances such as the methodologies established by Knechtle et al. often rely on Mean Absolute Error (MAE) to measure absolute deviation [24]. However, this study encompasses a highly heterogeneous dataset with diverse race distances and varying terrain difficulties. Consequently, Mean Absolute Percentage Error (MAPE) was prioritized as the primary performance indicator [22]. Unlike MAE, which is heavily influenced by the absolute magnitude of race durations, MAPE standardizes the prediction error as a percentage, providing a fair, scale-independent evaluation across events of varying lengths.

Table 1. Performance Comparison of Predictive Models (90:10 Split)

Model	MAE	RMSE	R2	MAPE (%)
Model A	0.9392	5.5951	0.7836	8.5047
Model B	1.0588	5.7434	0.7720	10.0490

As shown in Table 1, Model A achieved the highest predictive accuracy under realistic validation conditions, explaining 78.36% of the variance ($R^2 = 0.7836$) [24] with a MAPE of 8.50% [25]. Excluding weather variables (Model B) increased the MAPE to 10.05% and reduced the R^2 to 0.7720. To rigorously confirm whether the integration of environmental features yields a statistically significant improvement across validation cycles, a paired t-test was conducted across the five cross-validation folds. The detailed statistical breakdown is presented in Table 2.

Table 2. Paired t-Test Results Across Cross-Validation Folds

Metric	Mean (Model A)	Mean (Model B)	t-statistic	p-value	Significance (α=0.05)
MAE	0.9471	1.095	-26.389	< 0.001	Yes
RMSE	2.0617	2.3083	-4.296	0.013	Yes
R^2	0.9629	0.9535	4.066	0.015	Yes
MAPE (%)	9.1394	10.7014	-86.302	< 0.001	Yes

As detailed in Table 2, the results indicate statistically significant improvements across all evaluation metrics. Model A achieved significantly lower MAE ($p < 0.001$), RMSE ($p = 0.013$), and MAPE ($p < 0.001$), while also obtaining a significantly higher R^2 score ($p = 0.015$). It is worth noting the minor numerical variance between the cross-validation metrics in Table 2 and the final test metrics in Table 1; this variance is expected, as Table 2 reflects the aggregated validation performance across training folds during cross-validation, whereas Table 1 demonstrates the final model's generalization capability when evaluated against a completely independent test set.

The high absolute t-statistics for MAE (-26.389) and MAPE (-86.302) provide statistically significant evidence that the inclusion of environmental variables consistently and heavily minimizes prediction deviations rather than altering the data by random chance. These findings strongly indicate that weather variables act as meaningful contextual predictors in event-level prediction [8], [9]. While the inclusion of weather variables yields statistically significant improvements, the absolute increase in R^2 remains modest. This is highly realistic in trail running, where significant variance is inherently driven by unrecorded human factors such as individual pacing strategies, technical running skills, nutrition management, and psychological resilience [9], [25]. Consequently, reaching an R^2 near 1.0 is unrealistic without individual athlete physiological data.

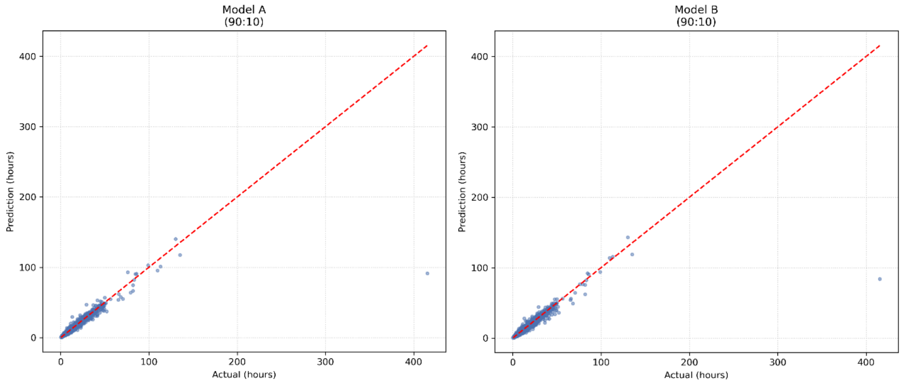

Figure 2. Parity plot demonstrating the relationship between actual and predicted Mean Finish Times (MFT) in hours for Model A and Model B, where the dashed red line represents the ideal $y = x$ reference line.

Figure 2 visualizes the prediction accuracy of Model A and B. The predicted values closely follow the ideal diagonal line ($y = x$), demonstrating high precision across a wide range of event durations, with only minor deviations at extreme ultra-endurance values (>100 hours). The analysis shows that RMSE values remain relatively high compared to MAE. This pattern is expected due to the squaring of residuals in RMSE, making it highly sensitive to extreme outliers [22].

In ultra-endurance events (e.g., >100 hours), large prediction deviations often occur due to unpredictable, non-environmental occurrences such as severe athlete fatigue, injuries, getting lost on the trail, or mid-race medical interventions [24], [26]. The substantial reduction in MAPE demonstrates that environmental information provides additional predictive signal beyond structural terrain characteristics alone. The effect becomes increasingly apparent in high *km_effort* races, where thermal and meteorological conditions appear to amplify overall event difficulty. These findings suggest that weather conditions help refine the relationship between terrain workload and predicted finish time, particularly in demanding ultra-endurance events.

3.3. Interpretability using SHAP

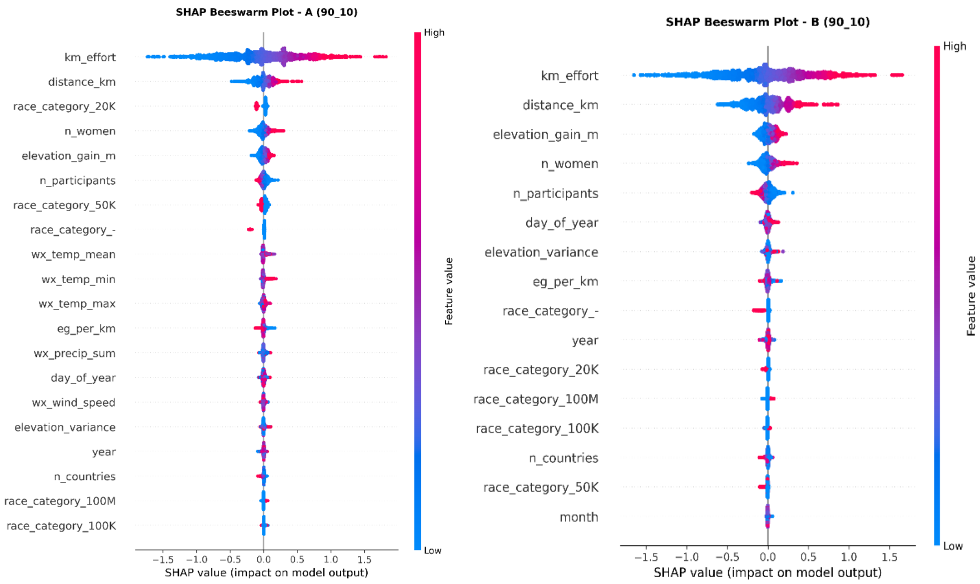

Figure 3. SHAP summary beeswarm plots for Model A and Model B, illustrating the distribution of feature impacts on the model's final prediction output, where red indicates high feature values and blue indicates low feature values.

As demonstrated by the reduction in MAE and MAPE (Table 1), integrating weather features significantly improves overall predictive accuracy. To mathematically understand how the model achieves this, Shapley Additive Explanations (SHAP) analysis (Figure 3) was utilized to reveal the specific feature attributions and internal decision-making process of the trained XGBoost model [12]. The beeswarm plot highlights that structural factors, namely *km_effort* and *distance_km*, remain the primary drivers of finish times, occupying the top positions with the widest distribution of SHAP values. Among the environmental variables, maximum temperature *(wx_temp_max)* and mean temperature *(wx_temp_mean)* serve as the strongest contextual modifiers.

Specifically, higher temperature values (red regions) correspond to positive SHAP contributions, indicating that elevated thermal conditions shift the model output toward longer predicted finish durations. Conversely, lower temperatures contribute negatively to the SHAP value distribution, resulting in shorter estimated race durations [8], [9].

However, it is important to emphasize that SHAP values describe the statistical behavior of the trained model rather than direct causal or physiological mechanisms in real-world endurance performance. Instead, they illustrate how environmental information is utilized by the model to refine finish time estimation under varying weather conditions.

3.4. Practical Implications for Race Management

The enhanced predictive accuracy achieved by integrating weather variables offers significant practical value for race organizers. Traditionally, Cut-Off Times (COT) and Cut-Off Points (COP) are established statically, relying almost entirely on fixed track distance and elevation gain as recommended by baseline guidelines [19]. However, this event-level model provides a data-driven foundation for calculating dynamic COTs and COPs that reflect actual environmental stress [8], [9]. By inputting forecasted weather conditions alongside structural terrain data, organizers can dynamically adjust time limits prior to the race. By incorporating forecasted weather conditions alongside terrain characteristics, organizers can dynamically adjust race time limits prior to the event. Such adaptation may improve participant safety during extreme weather exposure while also supporting more efficient allocation of medical and logistical resources [10], [27].

3.5. Limitations and Future Works

Despite achieving statistically significant predictive improvements, this study acknowledges several critical limitations. First, the macro-level daily weather data utilized cannot fully capture rapid microclimate variations such as altitude-dependent thermal drops, localized convective precipitation, or transient wind exposure frequently encountered along vast, topographically complex mountain routes [1], [28]. Second, the event-level framework deliberately excludes individual athlete variables (e.g., age, physiological capacity, nutrition, and pacing strategies), which inherently account for much of the unrecorded long-tail variance in ultra-endurance outcomes [29], [30]. Finally, reliance on the UTMB World Series ecosystem [16] introduces a geographic bias toward European alpine terrains and temperate climates, potentially limiting the model's direct generalizability to tropical, arid, or highly humid environments.

To address these boundaries, future research should focus on three main directions:

1. Integrating checkpoint-level, high-resolution sequential weather data to model real-time environmental variations along the race course.
2. Incorporating aggregated athlete demographics per event, such as mean age and gender distribution, to better capture collective participant capability under environmental stress [3], [29].
3. Validating the framework against independent external datasets, such as the ITRA Global database [19] or regional Trail Run Series, to verify its cross-ecosystem robustness and broader operational applicability.

4. CONCLUSION

This study demonstrates that incorporating environmental variables improves predictive reliability in event-level trail running finish time estimation. Using a leakage-controlled Group Shuffle Split evaluation, the proposed framework consistently reduced prediction error and achieved statistically significant improvements across all evaluation metrics. The findings indicate that weather conditions provide additional contextual information that helps refine the relationship between terrain difficulty and predicted finish duration, particularly in high-effort races. From a practical perspective, the proposed framework offers a data-driven basis for dynamically adjusting Cut-Off Times (COT) and Cut-Off Points (COP) to support safer and more adaptive race management under varying environmental conditions.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, for continuous academic support. The authors would also like to acknowledge Mr. Amin Hariyadi and all lecturers involved for their invaluable guidance, insightful feedback, and academic supervision throughout this research. The conceptual framework of balancing complex data variables in this study reflects the universal principle of precision and measure in natural systems.

REFERENCES

[1] L. P. Ardigò, C. Capelli, and G. P. Millet, "Editorial: Human Ultra-Endurance Exercise," Jun. 25, 2020, *Frontiers Media S.A.* doi: 10.3389/fphys.2020.00664.

[2] P. Belinchón-Demiguel, P. Ruisoto, B. Knechtle, P. T. Nikolaidis, B. Herrera-Tapias, and V. J. Clemente-Suárez, "Predictors of athlete's performance in ultra-endurance mountain races," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, pp. 1–8, Feb. 2021, doi: 10.3390/ijerph18030956.

[3] B. Knechtle *et al.*, "The fastest 24-hour ultramarathoners are from Eastern Europe," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-75260-0.

[4] J. Turnwald *et al.*, "Analysis of the 50-mile ultramarathon distance using a predictive XGBoost model," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-92581-w.

[5] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *in Advances in Neural Information Processing Systems (NeurIPS)*, 2017, [Online]. Available: https://github.com/Microsoft/LightGBM.

[6] R. Shwartz-Ziv and A. Armon, "Tabular Data: Deep Learning is Not All You Need," *Information Fusion*, Nov. 2021, doi: https://doi.org/10.1016/j.inffus.2021.11.013.

[7] D. Shu *et al.*, "Prediction of half-marathon performance of male recreational marathon runners using nomogram," *BMC Sports Sci. Med. Rehabil.*, vol. 16, no. 1, Dec. 2024, doi: 10.1186/s13102-024-00889-3.

[8] N. El Helou *et al.*, "Impact of Environmental Parameters on Marathon Running Performance," *PLoS One*, vol. 7, no. 5, p. e37407, May 2012, doi: 10.1371/JOURNAL.PONE.0037407.

[9] K. Mantzios *et al.*, "Effects of Weather Parameters on Endurance Running Performance: Discipline-specific Analysis of 1258 Races," *Med. Sci. Sports Exerc.*, vol. 54, no. 1, pp. 153–161, Jan. 2022, doi: 10.1249/MSS.0000000000002769.

[10] D. Gregory, K. Kintziger, S. Crouter, C. Sims, M. Kellogg, and E. Fitzhugh, "Weather effects on natural surface trail use in an urban wilderness multi-use trail system," *Journal of Outdoor Recreation and Tourism*, vol. 46, p. 100757, Jun. 2024, doi: 10.1016/J.JORT.2024.100757.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[12] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[13] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. R. Müller, "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 247–278, Mar. 2021, doi: 10.1109/JPROC.2021.3060483.

[14] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013. doi: 10.1007/978-1-4614-6849-3.

[15] "Mannasd/prediksi_mean_finish_time_trail_run: Implementasi pipeline machine learning untuk prediksi performa lari menggunakan data UTMB yang diperkaya dengan fitur cuaca dan variabel aktivitas." Accessed: May 05, 2026. [Online]. Available: https://github.com/Mannasd/prediksi_mean_finish_time_trail_run

[16] "UTMB World Series - Meet your extraordinary!" Accessed: May 05, 2026. [Online]. Available: https://utmb.world/

[17] " Free Open-Source Weather API | Open-Meteo.com." Accessed: May 05, 2026. [Online]. Available: https://open-meteo.com/

[18] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: http://scikit-learn.sourceforge.net.

[19] "ITRA - International Trail Running Association - official world ranking of trail runners recognized by World Athletics." Accessed: May 05, 2026. [Online]. Available: https://itra.run/

[20] H. Wang, N. Lu, T. Chen, H. He, Y. Lu, and X. M. Tu, "Log-transformation and its implications for data analysis," *Shanghai Arch. Psychiatry*, vol. 26, no. 2, pp. 105–109, 2014, doi: 10.3969/j.issn.1002-0829.2014.02.009.

[21] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[22] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.

[23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *in Advances Neural Information Processing Systems (NeurIPS)*, Jan. 2019, [Online]. Available: http://arxiv.org/abs/1706.09516

[24] M. Thuany *et al.*, "An analysis of the 6-h ultra-marathon race using a machine learning approach," *Front. Sports Act. Living*, vol. 7, 2025, doi: 10.3389/fspor.2025.1577470.

[25] E. W. Solang, L. Linawati, I. B. G. Manuaba, and I. N. Setiawan, "A Systematic Literature Review of Machine Learning for Endurance Running Performance Prediction," *sinkron*, vol. 10, no. 1, pp. 512–524, Jan. 2026, doi: 10.33395/sinkron.v10i1.15743.

[26] B. Knechtle *et al.*, "Race course characteristics are the most important predictors in 48 h ultramarathon running," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-94402-6.

[27] E. Y. Lee *et al.*, "Ambient environmental conditions and active outdoor play in the context of climate change: A systematic review and meta-synthesis," *Environ. Res.*, vol. 283, Oct. 2025, doi: 10.1016/j.envres.2025.122146.

[28] B. Knechtle *et al.*, "Change in elevation predicts 100 km ultra marathon performance," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-09502-0.

[29] M. J. Joyner, "Physiological limits to endurance exercise performance: influence of sex," May 01, 2017, *Blackwell Publishing Ltd*. doi: 10.1113/JP272268.

[30] S. Lazzer, D. Salvadego, E. Rejc, A. Buglione, G. Antonutto, and P. E. Di Prampero, "The energetics of ultra-endurance running," *European Journal of Applied Physiology 2011 112:5*, vol. 112, no. 5, pp. 1709–1715, Sep. 2011, doi: 10.1007/S00421-011-2120-Z.