PREDIKSI RATA-RATA WAKTU FINIS PADA LOMBA *TRAIL RUNNING* MENGGUNAKAN XGBOOST DAN *EXPLAINABLE AI*

SKRIPSI

Oleh:
MUHAMMAD ANNAS DARUNNAJA
NIM. 220605110130

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

PREDIKSI RATA-RATA WAKTU FINIS PADA LOMBA *TRAIL RUNNING* MENGGUNAKAN XGBOOST DAN *EXPLAINABLE AI*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
MUHAMMAD ANNAS DARUNNAJA
NIM. 220605110130

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

HALAMAN PERSETUJUAN

PREDIKSI RATA-RATA WAKTU FINIS PADA LOMBA *TRAIL RUNNING* MENGGUNAKAN XGBOOST DAN *EXPLAINABLE AI*

SKRIPSI

Oleh:
MUHAMMAD ANNAS DARUNNAJA
220605110130

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 12 Juni 2026

Pembimbing I,	Pembimbing II,
Dr. M. Amin Hariyadi, M.T	Okta Qomaruddin Aziz, M.Kom
NIP. 19670118 200501 1 001	NIP. 19911019 201903 1 013

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

PREDIKSI RATA-RATA WAKTU FINIS PADA LOMBA *TRAIL RUNNING* MENGGUNAKAN XGBOOST DAN *EXPLAINABLE AI*

SKRIPSI

Oleh:
MUHAMMAD ANNAS DARUNNAJA
NIM. 220605110130

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 23 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Johan Ericka Wahyu Prakasa, M.Kom NIP. 19831213 201903 1 004	( )
Anggota Penguji I	: Dr. Agung Teguh Wibowo Almais, M.T NIP. 19860301 202321 1 016	( )
Anggota Penguji II	: Dr. M. Amin Hariyadi, M.T NIP. 19670118 200501 1 001	( )
Anggota Penguji III	: Okta Qomaruddin Aziz, M.Kom NIP. 19911019 201903 1 013	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Muhammad Annas Darunnaja
NIM	: 220605110130
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Prediksi Rata-Rata Waktu Finis pada Lomba *Trail Running* Menggunakan XGBoost dan *Explainable AI*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilalihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Juni 2026
Yang membuat pernyataan,

Muhammad Annas Darunnaja
NIM.220605110130

MOTTO

"I don't have bad days. I just have days that are less good than others."
(Noel Gallagher)

"Maybe I just wanna fly. Wanna live, I don't wanna die"
(Oasis - Live Forever)

HALAMAN PERSEMBAHAN

Alhamdulillahi rabbil 'alamin, segala puji bagi Allah Swt. atas segala rahmat, nikmat, dan karunia-Nya. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad *Shallallahu 'Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh umatnya. Atas izin dan pertolongan Allah Swt., penulis diberikan kemudahan dan kekuatan hingga dapat menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada kedua orang tua tercinta sebagai wujud bakti, cinta, kasih sayang, dan rasa terima kasih atas segala doa, pengorbanan, serta dukungan yang telah diberikan. Kehadiran dan doa mereka menjadi alasan utama yang menguatkan penulis untuk terus berjuang hingga mampu menyelesaikan skripsi ini.

Kepada saudara-saudara tercinta dan seluruh keluarga besar, penulis mempersembahkan karya sederhana ini sebagai bentuk rasa syukur dan terima kasih atas perhatian, motivasi, dukungan, serta doa yang selalu menyertai perjalanan penulis dalam menempuh pendidikan hingga terselesaikannya skripsi ini.

Kepada teman-teman dan seluruh pihak yang telah turut membantu penulis dalam menyelesaikan skripsi ini, terima kasih atas doa, dukungan, dan segala bentuk bantuan yang diberikan sehingga skripsi ini dapat terselesaikan dengan baik.

KATA PENGANTAR

Alhamdulillāhirabbil ‘ālamīn, segala puji bagi Allah *Subḥānahu wa Ta‘ālā,* yang telah melimpahkan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Prediksi Rata-Rata Waktu Finis pada Lomba *Trail Running* Menggunakan XGBoost dan *Explainable AI*” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shallallahu ‘Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Bapak Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Dr. M. Amin Hariyadi, M.T., selaku dosen pembimbing I dan Bapak Okta Qomaruddin Aziz, M.Kom., selaku dosen pembimbing II juga Dosen wali, yang telah memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas

waktu, kesabaran, dan perhatian yang telah diberikan sehingga penelitian dan penulisan skripsi ini dapat terselesaikan dengan baik dan tepat waktu.

5. Bapak Johan Ericka Wahyu Prakasa, M.Kom. dan Dr. Agung Teguh Wibowo Almais, M.T., selaku Dosen Penguji yang telah memberikan kritik, saran, dan bimbingan untuk menyempurnakan skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, penulis menyampaikan ucapan terima kasih atas ilmu yang telah diberikan selama masa perkuliahan. Semoga segala ilmu dan kebaikan yang telah diberikan menjadi amal jariyah serta memberikan manfaat yang berkelanjutan bagi penulis di masa yang akan datang.
7. Ibu Nia Faricha, S.Si., admin Program Studi Teknik Informatika, yang membantu dalam urusan administrasi dan selalu mengingatkan tentang kelengkapan berkas.
8. Keluarga tercinta, khususnya Bapak, Ibu, kakak, dan adik penulis, yang selalu menjadi sumber motivasi terbesar melalui lantunan doa, limpahan kasih sayang, serta dukungan moril maupun materiel yang tiada putusnya.
9. Keluarga besar Teknik Informatika angkatan 2022 Infinity, yang telah memberikan kebersamaan, dukungan, dan pengalaman berharga selama masa perkuliahan.
10. Sahabat seperjuangan yang senantiasa menemani penulis selama masa perkuliahan, yaitu Ki Ageng Nasrokh Mangkunegoro Putra, Ahmad Fadhli Dzilikrom, Muhammad Andrean Hidayatulloh, Muhammad Alif Athallah,

Muhammad Zidan Alvaro, Muhammad Haikal Azadin, dan Rizky Fadilah Akbar. Penulis juga tidak lupa mengucapkan terima kasih kepada sahabat lainnya yang tergabung dalam grup ”Takmir Mastar”, yaitu Nizam Hukmul Kautsar, Ega Okta Syafan Prayoga, Akbar Bimantara Trahestiawan, Nur Muhammad Najiburrohman, Hasyim Husein Alhabsji, Radifan Roihanul Fiqri, dan Ahmad Zanuar Dito Ananda.

11. Sahabat yang selalu membersamai penulis dari SMP, SMA sampai sekarang, Mohammad Sirojuddin Mahfudz, Mochammad Sofyan Haqiqi, Siska Putri Cahyarani, Dimas Setyo Pambudi, Irnandini Putri Imroatus Sholihah, Riza Akbar Firmansyah, Sony Bagas Sin Albir, Muhammad Arifin Alfianto, dan Rizal Aprianto. Terima kasih atas dukungannya yang telah membawa penulis sampai di titik ini.
12. Teman-teman semasa SMP, SMA yang telah memberikan banyak pengalaman, kebersamaan, dan kenangan berharga. Ucapan terima kasih juga penulis sampaikan kepada sahabat-sahabat dan rekan-rekan lainnya yang hingga saat ini masih menjalin hubungan baik dengan penulis maupun yang pernah hadir dalam perjalanan hidup penulis. Terima kasih atas doa, dukungan, motivasi, kebersamaan, serta segala bentuk kebaikan yang telah diberikan. Meskipun tidak dapat disebutkan satu per satu, setiap perhatian dan bantuan yang diberikan menjadi cahaya yang menerangi langkah penulis hingga mampu menyelesaikan skripsi ini.

13. Teman-teman KKN penulis di Desa Srigading, yaitu Muhammad Azzam Al Habib, Mochammad Charis Firismanda, Andre Fernando, dan Muhammad Mujibur Rochman.
14. Seluruh keluarga, teman, sahabat, dan kerabat penulis yang tidak dapat penulis sebutkan satu per satu, yang turut memberikan bantuan, semangat, dukungan, dan doa untuk penulis.
15. Tim kebanggaan penulis, Timnas Jerman dan Real Madrid yang akan juara lagi tahun ini, yang pertandingannya selalu menjadi pelipur lara, hiburan, dan penyemangat tersendiri di sela-sela penatnya pengerjaan skripsi

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat berbagai keterbatasan yang memerlukan penyempurnaan di masa mendatang. Namun demikian, penulis meyakini bahwa proses belajar tidak memiliki batas akhir, sehingga setiap kritik dan saran yang membangun akan menjadi bekal berharga untuk terus berkembang. Oleh karena itu, penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang membutuhkan sebagai referensi untuk penelitian selanjutnya.

Malang, 12 Juni 2026

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Darunnaja, Muhammad Annas. 2026. **Klasifikasi Prediksi Rata-Rata Waktu Finis Pada Lomba *Trail Running* Menggunakan XGBoost Dan *Explainable AI***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T. (II) Okta Qomaruddin Aziz, M.Kom.

Kata Kunci: *Extreme Gradient Boosting*, *Explainable AI*, SHAP, *Trail Running*, rata-rata waktu finis.

Trail Running merupakan olahraga ketahanan dengan karakteristik medan yang kompleks, sehingga rata-rata waktu finis (*mean finish time*) pada level *event* menjadi indikator penting untuk mengukur tingkat kesulitan lomba. Sebagian besar penelitian yang ada masih berfokus pada prediksi performa individu atlet, sementara pemodelan berbasis karakteristik *event* secara agregat masih sangat terbatas. Penelitian ini mengembangkan model regresi berbasis *Extreme Gradient Boosting* (XGBoost) untuk memprediksi *mean finish time* menggunakan dataset UTMB *Race Data Sheet* periode 2014–2024. Empat skenario model dengan kombinasi fitur yang berbeda dievaluasi pada tiga skenario pembagian data (90:10, 80:20, 70:30) menggunakan *k-Fold Cross-Validation* (k=5) dan dioptimasi melalui *RandomizedSearchCV*. Untuk mengidentifikasi faktor dominan yang memengaruhi prediksi, digunakan metode *Shapley Additive Explanations* (SHAP) sebagai pendekatan Explainable AI. Hasil evaluasi menunjukkan bahwa 4 model menghasilkan performa terbaik dengan MAPE 8,85%, MAE 0,9047 jam, dan R^2 0,9570. Model juga terbukti stabil terhadap variasi pembagian data. Analisis SHAP mengungkapkan bahwa fitur km_*effort*, *elevation_gain_m*, dan *distance_km* merupakan kontributor paling dominan, sedangkan fitur cuaca berperan sebagai variabel pendukung yang meningkatkan akurasi secara moderat. Penelitian ini menyimpulkan bahwa XGBoost efektif dalam memodelkan *mean finish time* pada level *event*, dan hasilnya dapat dimanfaatkan oleh penyelenggara lomba untuk perencanaan operasional maupun oleh peserta untuk penyusunan strategi *Pacing* dan nutrisi yang lebih realistis.

ABSTRACT

Darunnaja, Muhammad Annas. 2026. **Prediction of *Mean finish time* in *Trail Running* Races Using XGBoost and Explainable AI**. Undergraduate thesis. Department of Computer Science, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisors: (I) Dr. M. Amin Hariyadi, M.T. (II) Okta Qomaruddin Aziz, M.Kom.

Keywords: *Extreme Gradient Boosting (XGBoost), Explainable AI, SHAP, Trail Running, mean finish time*

Trail Running is an endurance sport characterized by complex terrain, varying elevation, and diverse environmental conditions, making mean finish time at the event level a critical indicator of race difficulty. However, most existing studies focus on individual athlete performance prediction, leaving event-level aggregate modeling largely underexplored in the literature. This study develops a regression model based on Extreme Gradient Boosting (XGBoost) to predict mean finish time using the UTMB Race Data Sheet dataset spanning 2014–2024. Four model scenarios with different feature combinations were evaluated across three data *split* configurations (90:10, 80:20, and 70:30) using k-Fold Cross-*Validation* (k=5) and optimized via RandomizedSearchCV. Shapley Additive Explanations (SHAP) was employed as an Explainable AI approach to identify the dominant features influencing predictions. Results show that some of model achieved the best overall performance with a MAPE of 8.85%, MAE of 0.9047 hours, and R^2 of 0.9570. The model also demonstrated strong Robustness across varying data *split* configurations. SHAP analysis revealed that km_effort, elevation_gain_m, and distance_km are the most dominant predictors, while weather-related features serve as secondary variables that moderately improve accuracy. This study concludes that XGBoost effectively models mean finish time at the event level, and the resulting system can support race organizers in operational planning and assist participants in developing more realistic Pacing and nutrition strategies.

مستخلص البحث

داروناجا، محمد أناس .2026 .توقع متوسط وقت الوصول إلى خط النهاية في سباقات الجري على الطرق الوعرة باستخدام XGBoostوالذكاء الاصطناعي القابل للتفسير. بحث جامعي. قسم هندسة المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفون: (1) د. م. أمين هاريادي، الماجستير، (2) أوكتا قمر الدين عزيز، الماجستير.

الكلمات المفتاحية: تعزيز التدرج المتطرف (XGBoost)، الذكاء الاصطناعي القابل للتفسير، SHAP، الجري على الطرق الوعرة، متوسط وقت الوصول إلى خط النهاية.

الجري على الطرق الوعرة هو رياضة تحمل تتميز بتضاريس معقدة، وارتفاعات متفاوتة، وظروف بيئية متنوعة، مما يجعل متوسط وقت الوصول إلى خط النهاية على مستوى الحدث مؤشرًا حاسمًا على صعوبة السباق. ومع ذلك، تركز معظم الدراسات الحالية على التنبؤ بأداء الرياضيين الفردي، مما يترك النمذجة الإجمالية على مستوى الحدث غير مستكشفة إلى حد كبير في الأدبيات. تطور هذه الدراسة نموذج انحدار قائم على تقنية تعزيز التدرج المتطرف ((XGBoostللتنبؤ بمتوسط وقت الوصول إلى خط النهاية باستخدام مجموعة بيانات UTMB Race Data Sheet التي تغطي الفترة من 2014 إلى 2024. تم تقييم أربعة سيناريوهات للنموذج مع تركيبات مختلفة من السمات عبر ثلاثة تكوينات لتقسيم البيانات (90:10، و80:20، و70:30) باستخدام التحقق المتقاطع k-fold (k=5)وتم تحسينها عبر *RandomizedSearchCV*. تم استخدام تفسيرات شابلي الإضافية (SHAP)كنهج للذكاء الاصطناعي القابل للتفسير لتحديد السمات المهيمنة التي تؤثر على التنبؤات. أظهرت النتائج أن النموذج ذو الميزات الكاملة (النموذج أ) حقق أفضل أداء إجمالي بمعدل الخطأ المتوسط المئوي (MAPE) قدره 8.85٪، ومعدل الخطأ المتوسط ((MAEقدره 0.9047 ساعة، ومعامل R^2 قدره 0.9570. كما أظهر النموذج متانة قوية عبر تكوينات تقسيم البيانات المتنوعة. وكشف تحليل SHAP أن المتغيرات *km_effort* و *elevation_gain_m* و *distance_km* هي المتغيرات التنبؤية الأكثر تأثيرًا، في حين تعمل السمات المتعلقة بالطقس كمتغيرات ثانوية تعمل على تحسين الدقة بشكل معتدل. وتخلص هذه الدراسة إلى أن XGBoost ينمذج بشكل فعال متوسط وقت الوصول إلى خط النهاية على مستوى الحدث، ويمكن للنظام الناتج أن يدعم منظمي السباق في التخطيط التشغيلي ويساعد المشاركين في تطوير استراتيجيات أكثر واقعية فيما يتعلق بوتيرة الجري والتغذية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Trail Running merupakan cabang olahraga ketahanan (*endurance sport*) yang dilakukan pada lintasan alam dengan karakteristik medan yang beragam, meliputi variasi elevasi, jenis permukaan lintasan, dan kondisi iklim (Gregory *et al.,* 2024). Kombinasi karakteristik tersebut menyebabkan durasi penyelesaian lomba dapat berbeda secara signifikan meskipun jarak tempuhnya identik. Dalam konteks penyelenggaraan dan perencanaan *event*, rata-rata waktu penyelesaian lomba (*mean finish time*) dapat diperlakukan sebagai indikator kuantitatif yang relatif objektif untuk merepresentasikan kesulitan sebuah *event* pada skala makro (Belinchón-Demiguel *et al.,* 2021).

Seiring meningkatnya partisipasi dan profesionalisasi penyelenggaraan *Trail Running* secara global, kebutuhan pemanfaatan data historis untuk evaluasi tingkat kesulitan dan perencanaan lomba menjadi semakin penting. Laporan *Sports & Fitness Industry Association (SFIA)* tahun 2024 mencatat peningkatan partisipasi *Trail Running* dari 13,2 juta pada tahun 2022 menjadi 14,8 juta pada tahun 2023, dengan pertumbuhan kumulatif pascapandemi sebesar 25,6%. Kondisi tersebut mengindikasikan bahwa kebutuhan informasi yang lebih terukur terkait durasi lomba dan tuntutan *event* juga meningkat, baik bagi pelari maupun penyelenggara.

Kebutuhan tersebut semakin nyata pada pelari pemula maupun atlet yang melakukan peningkatan kategori lomba, misalnya dari jarak pendek ke jarak lebih panjang atau dari lintasan relatif datar ke lintasan dengan profil elevasi lebih

kompleks. Kelompok ini kerap mengalami kesulitan dalam memilih *event* yang sesuai serta memperkirakan besaran *effort* yang dibutuhkan. Ketidakakuratan estimasi durasi dan tuntutan fisiologis berpotensi meningkatkan risiko gangguan kesehatan seperti deplesi glikogen, dehidrasi, dan hiponatremia, sebagaimana ditekankan dalam panduan *Nutrition for Ultramarathon Running* (Costa *et al.,* 2019). Selain itu, beberapa studi menunjukkan bahwa keterbatasan dalam memperkirakan durasi dan tuntutan fisiologis lomba dapat berkontribusi terhadap kesalahan strategi *Pacing*, kelelahan dini, serta meningkatnya risiko gagal menyelesaikan lomba, khususnya pada pelari dengan pengalaman terbatas di medan *Trail Running* (Suter *et al.,* 2020).

Namun, praktiknya saat ini estimasi waktu penyelesaian masih sering mengandalkan pengalaman subjektif atau pendekatan statistik sederhana yang berfokus pada karakteristik individu pelari (*athlete-based*), seperti usia atau indeks massa tubuh, dengan asumsi hubungan *linear*. Pada *event Trail Running*, asumsi tersebut berpotensi tidak memadai karena durasi penyelesaian dapat berubah secara substansial akibat interaksi *on-linear* dan multivariat antara karakteristik lintasan, kondisi lingkungan, serta skala *event.* Oleh karena itu, pendekatan berbasis karakteristik *event* (*event-based*) dinilai lebih relevan untuk mendukung kebutuhan perencanaan pada tingkat penyelenggaraan (Knechtle *et al.,* 2024).

Dalam olahraga *endurance,* durasi penyelesaian lomba dipandang sebagai hasil interaksi multivariat yang bersifat *non-linear* karakteristik lintasan dan kondisi lingkungan. Meskipun berbagai penelitian telah memanfaatkan pendekatan statistik maupun *Machine learning* untuk memodelkan performa pada olahraga daya tahan,

sebagian besar penelitian tersebut masih berfokus pada prediksi performa individu atlet (*Athlete-based prediction)* yang menggunakan variabel fisiologis atau riwayat performa pelari sebagai prediktor utama. Pendekatan tersebut memiliki keterbatasan pada analisis tingkat kesulitan *event* secara agregat, karena variasi durasi lomba pada *Trail Running* tidak hanya dipengaruhi oleh karakteristik individu, tetapi juga oleh kombinasi karakteristik lintasan dan kondisi lingkungan. Selain itu, integrasi variabel lintasan dan faktor lingkungan historis dalam model prediksi durasi lomba masih relatif terbatas dalam literatur. Oleh karena itu, diperlukan penelitian yang secara khusus memodelkan *mean finish time* pada level *general event* menggunakan pendekatan *Machine learning* yang mampu menangkap hubungan *non-linear* antar variabel.

Dalam penelitian ini, *Extreme Gradient Boosting* (XGBoost) dipilih sebagai algoritma utama karena memiliki beberapa keunggulan dalam pemodelan data tabular yang kompleks. XGBoost merupakan metode *ensemble* berbasis *gradient boosting* yang membangun model secara iteratif dengan meminimalkan kesalahan residual pada setiap tahap pembelajaran. Dibandingkan pendekatan regresi linier konvensional, XGBoost mampu menangkap hubungan *non-linear* serta interaksi antar variabel tanpa bergantung pada asumsi distribusi tertentu. Selain itu, algoritma ini dilengkapi mekanisme regularisasi L1 dan L2 yang membantu mengurangi risiko *overfitting* serta meningkatkan kemampuan generalisasi model pada data baru (Chen & Guestrin, 2016). Sejumlah penelitian pada bidang olahraga daya tahan juga menunjukkan bahwa algoritma *ensemble* seperti XGBoost mampu menghasilkan performa prediktif yang lebih baik dibandingkan metode statistik

tradisional pada data tabular dengan variabel heterogen (Solang *et al.,* 2026; Turnwald *et al.,* 2025).

Penelitian ini menggunakan dataset *event Trail Running* dari UTMB Race *Data Sheet* periode 2014–2024 dari *repository* Kaggle. Fokus penelitian diarahkan pada pengembangan model prediksi *mean finish time* pada tingkat *event* serta evaluasi kinerja model XGBoost sebelum dan sesudah *hyperparameter Tuning* untuk menilai peningkatan akurasi dan reliabilitas prediksi. Selain aspek akurasi, penelitian ini menekankan kebutuhan interpretabilitas agar faktor dominan yang memengaruhi durasi *event* dapat diidentifikasi secara transparan. Oleh karena itu, penelitian mengintegrasikan *Shapley Additive Explanations* (SHAP) untuk mengukur kontribusi relatif variabel prediktor terhadap estimasi durasi lomba (Lundberg *et al.,* 2020*).* Dengan demikian, penelitian tidak hanya menghasilkan prediksi, tetapi juga menghasilkan penjelasan yang mendukung pengambilan keputusan berbasis data pada level *event*

Selain pertimbangan ilmiah, upaya analisis dan prediksi dalam penelitian ini juga dipandang selaras dengan nilai keislaman yang menekankan pentingnya perencanaan dan evaluasi dalam ikhtiar manusia. Allah SWT berfirman dalam Q.S. Al-Hasyr ayat 18:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوا اتَّقُوا اللّٰهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍۚ وَاتَّقُوا اللّٰهَۗ اِنَّ اللّٰهَ خَبِيْرٌۢ بِمَا تَعْمَلُوْنَ

“Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Dan bertakwalah kepada Allah. Sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.” (Q.S Al-Hasyr : 18)

Menurut Tafsir Ibnu Katsir, Q.S. Al-Hasyr ayat 18 memuat perintah untuk melakukan *muhasabah* (evaluasi diri) serta mempersiapkan hal-hal yang akan dihadapi di masa mendatang secara sadar dan bertanggung jawab (Tafsir Ibnu Katsir, 1923). Nilai perencanaan, kehati-hatian, dan tanggung jawab tersebut sejalan dengan prinsip ilmiah dalam melakukan evaluasi dan prediksi berbasis data yang terukur.

Oleh karena itu, penelitian ini tidak hanya berorientasi pada pengembangan model prediktif, tetapi juga memberikan kontribusi praktis pada dua level utama, yaitu penyelenggara *event* dan peserta lomba. Pada level penyelenggara, estimasi *mean finish time* berbasis karakteristik lintasan dan kondisi lingkungan dapat digunakan sebagai dasar perencanaan *cut-off time*, strategi distribusi logistik (hidrasi, nutrisi, dan *medical support*), serta manajemen operasional lomba secara lebih presisi dan berbasis data. Sementara itu, pada level peserta, hasil prediksi dan interpretasi faktor dominan yang memengaruhi durasi *event* dapat menjadi referensi objektif dalam memilih kategori lomba yang sesuai, merencanakan strategi *Pacing*, serta menyusun kebutuhan nutrisi dan hidrasi secara lebih realistis.

Penelitian ini diharapkan tidak hanya memperkaya kajian akademik dalam bidang analitik olahraga ketahanan berbasis *ensemble learning*, tetapi juga menghadirkan manfaat aplikatif yang konkret dalam mendukung keselamatan, efektivitas perencanaan, dan pengambilan keputusan berbasis data pada ekosistem *Trail Running*.

1.2 Pernyataan Masalah

Berdasarkan latar belakang penelitian, identifikasi masalah dirumuskan sebagai berikut:

1. Dibutuhkan pemodelan prediktif pada level *event* yang mampu memperkirakan *mean finish time* secara terukur menggunakan kombinasi karakteristik lintasan dan kondisi lingkungan/cuaca.
2. Belum tersedia analisis interpretabilitas yang transparan untuk menjelaskan kontribusi relatif tiap fitur lomba *Trail Running* terhadap hasil prediksi *mean finish time*, sehingga faktor dominan yang memengaruhi durasi *event* belum dapat diidentifikasi secara jelas untuk mendukung pengambilan keputusan berbasis data.

1.3 Batasan Masalah

Untuk menjaga fokus dan keterarahan penelitian, batasan masalah dirumuskan sebagai berikut:

1. Penelitian menggunakan data sekunder *event Trail Running* dari UTMB Race Data Sheet periode 2014–2024 yang tersedia melalui platform Kaggle dengan unit analisis pada tingkat *event*.
2. Variabel penelitian dibatasi pada karakteristik lintasan dan kondisi lingkungan yang tersedia pada dataset penelitian.
3. Model belum memperhitungkan teknikalitas permukaan tanah.
4. Penelitian tidak memodelkan performa individu atlet maupun dinamika lomba secara *real-time*.

5. Penelitian berfokus pada pengembangan dan evaluasi model XGBoost serta interpretasi kontribusi fitur menggunakan SHAP.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mengembangkan model regresi berbasis XGBoost untuk memprediksi *mean finish time* pada *event Trail Running* berdasarkan karakteristik lintasan dan kondisi lingkungan/cuaca.
2. Menganalisis dan menginterpretasikan kontribusi masing-masing variabel prediktor terhadap hasil prediksi menggunakan SHAP, guna mengidentifikasi faktor dominan yang memengaruhi *mean finish time* pada level *event*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan menghasilkan estimasi *mean finish time* yang lebih terukur sebagai dasar objektif dalam perencanaan *event Trail Running*. Hasil penelitian dapat dimanfaatkan untuk mendukung perencanaan nutrisi, hidrasi, dan strategi *Pacing* yang realistis, serta membantu penyelenggara lomba dalam pengelolaan waktu operasional dan logistik *event*. Selain itu, penelitian ini berkontribusi pada pengembangan analitik olahraga ketahanan melalui penerapan *ensemble learning* dan pendekatan *explainable learning* pada konteks prediksi durasi lomba berbasis karakteristik *event*.

BAB II

STUDI PUSTAKA

Bab ini membahas landasan teori yang relevan dengan penelitian serta penelitian-penelitian terdahulu yang berkaitan dengan topik yang dikaji. Studi pustaka disusun untuk memperkuat dasar konseptual penelitian, memberikan gambaran mengenai pendekatan dan temuan penelitian sebelumnya, serta mengidentifikasi celah penelitian (*research gap*) yang menjadi dasar pengembangan penelitian ini.

2.1 Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai acuan dalam mengembangkan penelitian yang dilakukan saat ini. Melalui kajian terhadap penelitian sebelumnya, peneliti dapat memahami karakteristik objek yang diteliti, variabel yang berperan, serta pendekatan numerik dan komputasional yang digunakan dalam menganalisis performa olahraga daya tahan. Penelitian-penelitian terdahulu dalam subbab ini diuraikan secara runtut sesuai dengan urutan pada Tabel 2.1.

Penelitian oleh Ardigò *et al.,* (2020) mengkaji *Trail Running* sebagai olahraga daya tahan dengan karakteristik lintasan yang kompleks dan dinamis. Melalui analisis kuantitatif terhadap variabel fisiologis serta karakteristik medan, penelitian tersebut menunjukkan bahwa variasi elevasi dan teknikalitas lintasan meningkatkan beban fisiologis pelari secara signifikan. Temuan ini mendukung penggunaan durasi penyelesaian lomba sebagai indikator representatif performa

dalam konteks *Trail Running,* sehingga relevan dengan pemilihan *mean finish time* sebagai variabel target dalam penelitian ini.

Belinchón-Demiguel *et al.,* (2021) menganalisis performa pelari pada berbagai *event Trail Running* menggunakan pendekatan statistik multivariat. Hasil penelitian menunjukkan bahwa perbedaan karakteristik lintasan, khususnya jarak dan elevasi, menyebabkan variasi waktu penyelesaian lomba yang signifikan antar *event* meskipun memiliki jarak tempuh yang serupa. Temuan tersebut memperkuat pendekatan berbasis *event* dalam memahami variasi performa, serta mendukung pemodelan *mean finish time* sebagai representasi tingkat kesulitan lintasan pada level *event*.

Pendekatan numerik yang lebih terstruktur ditunjukkan dalam penelitian Jaszczak dan Płociniczak (2024), yang mengembangkan model matematis dinamis untuk menganalisis strategi optimal pada *Trail Running* jarak jauh. Dengan mengintegrasikan faktor lintasan, nutrisi, dan dinamika kelelahan (fatigue), serta menerapkan teori optimal control melalui Pontryagin Maximum Principle, penelitian tersebut menghasilkan estimasi waktu penyelesaian lomba dengan galat relatif kecil. Studi ini menegaskan bahwa durasi lomba merupakan hasil interaksi multivariat yang kompleks, sehingga pendekatan pemodelan *non-linear* lebih sesuai dibandingkan model linier sederhana.

Pengaruh faktor lingkungan terhadap waktu penyelesaian lomba dikaji oleh Wilson dan Wilson (2024) melalui pendekatan Bayesian Linear Mixed Model. Penelitian tersebut menunjukkan bahwa kondisi cuaca, khususnya curah hujan pada bulan lomba maupun periode sebelumnya, berkontribusi terhadap peningkatan

waktu finish pada lomba *cross country*, sementara jarak tetap menjadi faktor paling konsisten dalam menentukan durasi lomba. Temuan ini mengindikasikan bahwa variabel lingkungan perlu diintegrasikan secara eksplisit dalam pemodelan durasi lomba, terutama pada *event* dengan karakteristik medan alam terbuka seperti *Trail Running*.

Dalam konteks *Machine learning*, Turnwald *et al.,* (2025) memodelkan performa pelari pada lomba *ultramarathon 50-mile* dengan memanfaatkan data elevasi, perubahan ketinggian, *split time*, serta kondisi lingkungan. Penelitian tersebut menggunakan algoritma XGBoost untuk memprediksi kecepatan dan waktu tempuh individu. Hasilnya menunjukkan bahwa XGBoost mampu menangkap hubungan *non-linear* antar variabel secara efektif, dengan elevasi lintasan menjadi salah satu kontributor utama terhadap variasi performa. Keberhasilan ini menunjukkan bahwa algoritma *ensemble* berbasis boosting efektif dalam konteks data endurance dengan karakteristik kompleks. Namun demikian, penelitian tersebut berfokus pada prediksi performa individu (*Athlete-based prediction*), berbeda dengan penelitian ini yang memodelkan *mean finish time* pada level *event* sebagai indikator agregat tingkat kesulitan lintasan.

Kajian komprehensif mengenai penerapan *Machine learning* dalam prediksi performa lari daya tahan dilakukan melalui *systematic literature review* oleh Solang *et al.,* (2026). Tinjauan terhadap penelitian periode 2020–2025 tersebut menunjukkan bahwa algoritma *ensemble learning* seperti *Random Forest,* XGBoost, dan LightGBM secara konsisten menghasilkan performa prediktif yang lebih baik dibandingkan regresi linier pada data *endurance* dan *ultramarathon*.

XGBoost dinilai memiliki stabilitas tinggi pada data tabular dengan variabel heterogen serta mampu menangkap interaksi *non-linear* antar fitur secara efisien. Temuan ini memberikan dasar metodologis yang kuat dalam pemilihan XGBoost sebagai algoritma utama dalam penelitian ini. Namun demikian, literatur yang dirangkum dalam kajian tersebut masih didominasi oleh pendekatan *athlete-based*, sementara pemodelan berbasis karakteristik *event* secara agregat masih relatif terbatas.

Selain aspek akurasi prediksi, interpretabilitas model juga menjadi perhatian penting dalam penelitian berbasis *Machine learning*. Molnar *et al.,* (2020) menekankan pentingnya *Explainable Artificial Intelligence* (*XAI)* dalam meningkatkan transparansi model prediktif, khususnya model *non-linear* dan *ensemble* berbasis pohon. Metode SHAP (*Shapley Additive Explanations)* memungkinkan analisis kontribusi masing-masing variabel terhadap hasil prediksi secara lokal maupun global, sehingga model yang kompleks tetap dapat dijelaskan secara ilmiah. Pendekatan ini relevan dalam penelitian ini untuk mengidentifikasi faktor lintasan dan lingkungan yang paling dominan memengaruhi *mean finish time* pada level *event*.

Ringkasan mengenai temuan utama dari penelitian-penelitian terdahulu yang relevan dengan penelitian ini disajikan pada Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Fokus Studi	Unit Analisis	Metode	Kontribusi Utama	Keterbatasan
1	Ardigò *et al.,* 2020	Pengaruh elevasi & medan terhadap beban fisiologis	Individu	Analisis statistik fisiologi olahraga	Elevasi dan teknikalitas meningkatkan kompleksitas beban fisiologis; durasi lomba	Tidak memodelkan prediksi berbasis data; tidak berbasis *event*-level

No	Peneliti	Fokus Studi	Unit Analisis	Metode	Kontribusi Utama	Keterbatasan
					representatif sebagai indikator performa	
2	Belinchón-Demiguel *et al.,* 2021	Variasi waktu finish antar *event*	*Event*	Analisis statistik multivariat	Profil lintasan menyebabkan variasi waktu finish signifikan antar *event*	Tidak menggunakan pendekatan prediktif *Machine learning*
3	Jaszczak & Płociniczak, 2024	Strategi optimal & estimasi waktu lomba	Individu	Model matematis dinamis & optimal control	Durasi lomba dipengaruhi interaksi elevasi–nutrisi–fatigue	Model teoritis; belum berbasis data historis *event*
4	Wilson & Wilson, 2024	Pengaruh cuaca terhadap waktu finish	*Event*	Bayesian Linear Mixed Model	Curah hujan & jarak memengaruhi durasi lomba	Model linier; belum menangkap *non-linear*itas kompleks
5	Turnwald *et al.,* 2025	Prediksi performa ultramarathon 50-mile	Individu	XGBoost	*Ensemble* boosting efektif menangkap hubungan *non-linear* pada data endurance	Fokus pada prediksi individu, bukan *mean finish time event*
6	Solang *et al.,* 2026	Tinjauan ML pada endurance running	Campuran (dominan individu)	Systematic Literature Review	*Ensemble* (RF, XGBoost, LightGBM) unggul pada data endurance tabular	Minim penelitian berbasis *event*-level agregat
7	Molnar *et al.,* 2020	Interpretabilitas model ML	Model	SHAP & XAI	Model *non-linear* dapat dijelaskan secara lokal & global	Tidak spesifik pada konteks *Trail Running*

Berdasarkan tinjauan literatur tersebut, terlihat bahwa sebagian besar penelitian masih berfokus pada prediksi performa individu atlet dan belum banyak yang menganalisis karakteristik *event* secara agregat. Selain itu, integrasi variabel lintasan dan faktor lingkungan historis dalam pemodelan prediksi durasi lomba masih relatif terbatas. Oleh karena itu, penelitian ini mengusulkan pendekatan prediktif berbasis *event* level menggunakan algoritma XGBoost yang dipadukan

dengan analisis interpretabilitas SHAP untuk mengidentifikasi faktor dominan yang memengaruhi *mean finish time* pada *event Trail Running*.

2.2 *Trail Running*

Trail Running merupakan cabang olahraga lari jarak jauh yang dilakukan pada lintasan alam terbuka seperti pegunungan, hutan, perbukitan, dan jalur *off-road*. Berbeda dengan *road running* yang umumnya berlangsung pada permukaan datar dan homogen, *Trail Running* memiliki karakteristik medan yang tidak rata, tingkat teknikalitas jalur yang bervariasi, serta perubahan elevasi yang signifikan. Karakteristik ini menyebabkan beban fisiologis dan mekanis yang dialami pelari menjadi lebih kompleks dibandingkan lari jalan raya (Ardigò *et al.,* 2020; Joyner, 2017).

Dalam *Trail Running*, performa pelari sangat dipengaruhi oleh kemampuan adaptasi terhadap medan, strategi *Pacing*, serta manajemen energi selama lomba. Variasi elevasi dan teknikalitas jalur menyebabkan perbedaan waktu penyelesaian lomba yang signifikan antar *event*, bahkan pada jarak yang sama. Oleh karena itu, *Trail Running* sering dipandang sebagai olahraga berbasis durasi (*time-based endurance sport*), sehingga waktu penyelesaian lomba menjadi indikator penting dalam evaluasi performa dan perencanaan strategi lomba (Belinchón-Demiguel *et al.,* 2021).

2.3 Organisasi dan Standardisasi *Trail Running* (ITRA & UTMB)

Perkembangan *Trail Running* secara global didukung oleh keberadaan lembaga internasional yang berperan dalam standarisasi *event* dan pengelolaan data lomba. *International Trail Running Association* (ITRA) merupakan organisasi yang

menetapkan klasifikasi jarak lomba, sistem penilaian tingkat kesulitan lintasan, serta indeks performa pelari dan *event*. Salah satu kontribusi utama ITRA adalah pengembangan parameter kesulitan lintasan berbasis jarak, elevasi, dan karakteristik teknis jalur yang digunakan secara luas dalam evaluasi *event Trail Running* (International *Trail Running* Association, 2022).

Selain ITRA, *UTMB World Series* berperan sebagai salah satu penyelenggara dan pengarsip kompetisi *Trail Running* internasional yang paling berpengaruh. UTMB tidak hanya menyelenggarakan lomba, tetapi juga menyediakan *race data sheet* yang terstruktur dan konsisten untuk berbagai *event* di seluruh dunia. Ketersediaan data terstandarisasi ini memungkinkan analisis performa berbasis *event* secara objektif dan komparatif, serta mendukung penerapan pendekatan analitik dan *Machine learning* dalam penelitian *Trail Running* (UTMB Group, 2023).

2.4 Faktor Lingkungan dalam *Trail Running*

Durasi lomba *Trail Running* sangat dipengaruhi oleh karakteristik lintasan dan kondisi lingkungan pada skala makro *event*. Faktor utama yang berperan meliputi jarak lintasan, total elevasi (*total ascent*), serta rasio elevasi per kilometer (*Elevation Gain* per Kilometer). Semakin besar elevasi yang harus ditempuh pelari dalam jarak tertentu, semakin tinggi beban fisiologis yang dialami, sehingga berdampak langsung pada waktu penyelesaian lomba (Knechtle *et al.,* 2024; Suter *et al.,* 2020).

Selain karakteristik lintasan, faktor lingkungan seperti suhu udara dan ketinggian lokasi lomba turut memengaruhi performa pelari. Suhu lingkungan yang

tinggi dapat meningkatkan risiko dehidrasi dan kelelahan, sedangkan ketinggian lokasi berkaitan dengan penurunan tekanan parsial oksigen yang berdampak pada kapasitas aerobik. Kombinasi faktor lintasan dan lingkungan ini membentuk kompleksitas *event Trail Running* yang menyebabkan variasi *mean finish time* antar *event* (Belinchón-Demiguel *et al.,* 2021)

2.5 Prediksi *Mean finish time* sebagai Masalah Regresi

Dalam konteks analisis performa *Trail Running*, *mean finish time* didefinisikan sebagai rata-rata waktu penyelesaian lomba seluruh peserta yang berhasil finis pada suatu *event*. Nilai ini merepresentasikan karakteristik durasi lomba secara keseluruhan dan dapat digunakan sebagai indikator objektif tingkat kesulitan lintasan pada level *event* (*event-based*) (Knechtle *et al.,* 2024).

Secara komputasional, prediksi *mean finish time* dikategorikan sebagai masalah regresi, karena variabel target bersifat numerik kontinu. Pendekatan regresi memungkinkan pemodelan hubungan antara *mean finish time* dan berbagai variabel prediktor seperti jarak, elevasi, dan kondisi lingkungan. Model regresi *non-linear* dinilai lebih sesuai untuk menangkap interaksi kompleks antar variabel dibandingkan pendekatan statistik linier konvensional (Lerebourg *et al.,* 2022; Shu *et al.,* 2024).

2.6 Algoritma *Ensemble Learning* dalam Prediksi Performa Olahraga

Ensemble learning merupakan pendekatan *Machine learning* yang mengombinasikan beberapa model prediktor secara iteratif untuk meningkatkan akurasi dan stabilitas hasil prediksi. Pendekatan ini sangat efektif dalam

memodelkan data tabular dengan hubungan *non-linear* dan interaksi kompleks antar variabel, seperti yang umum dijumpai pada analisis performa olahraga daya tahan.

Salah satu algoritma *ensemble learning* yang banyak digunakan dalam pemodelan prediktif modern adalah *Extreme Gradient Boosting* (XGBoost). XGBoost merupakan metode boosting berbasis pohon keputusan yang mengoptimalkan fungsi objektif secara bertahap dengan meminimalkan kesalahan prediksi residual pada setiap iterasi. Algoritma ini memiliki keunggulan dalam efisiensi komputasi, kemampuan regularisasi untuk mencegah *overfitting*, serta performa prediksi yang tinggi pada data tabular berdimensi menengah hingga besar.

Dalam konteks prediksi performa olahraga, XGBoost telah terbukti mampu menangkap hubungan *non-linear* antara karakteristik lintasan, faktor lingkungan, dan variabel durasi lomba secara efektif. Oleh karena itu, XGBoost dipilih sebagai algoritma utama dalam penelitian ini untuk memodelkan *mean finish time* pada *event Trail Running* berbasis karakteristik lintasan dan kondisi lingkungan (Chen & Guestrin, 2016; Turnwald *et al.,* 2025).

2.7 Explainable AI (*Shapley Additive Explanations*)

Model *Machine learning* kompleks seperti XGBoost sering kali dianggap sebagai *black-box* karena sulit diinterpretasikan secara langsung. Untuk meningkatkan transparansi dan validitas ilmiah, pendekatan *Explainable Artificial Intelligence* (XAI) dikembangkan guna menjelaskan mekanisme kerja model prediktif (Molnar *et al.,* 2020).

Salah satu metode XAI yang paling banyak digunakan adalah *Shapley Additive Explanations* (SHAP), yang didasarkan pada teori permainan (*game theory*) untuk mendistribusikan kontribusi setiap fitur terhadap prediksi akhir model. SHAP memungkinkan analisis kontribusi variabel secara lokal maupun global pada model *non-linear* dan *ensemble* berbasis pohon, sehingga membantu mengidentifikasi faktor dominan serta arah pengaruh variabel terhadap hasil prediksi (Lundberg *et al.,* 2020; Molnar *et al.,* 2020).

2.8 Evaluasi Model

Evaluasi kinerja model regresi bertujuan untuk menilai seberapa baik model memprediksi nilai aktual *mean finish time*. Dalam penelitian ini digunakan beberapa metrik evaluasi yang umum dalam masalah regresi, yaitu *Coefficient of Determination* (R^2), *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) (Chicco *et al.,* 2021).

Nilai R^2 mengukur proporsi variasi data yang dapat dijelaskan oleh model, sedangkan MAE dan RMSE mengukur besar kesalahan prediksi dalam satuan waktu. RMSE bersifat lebih sensitif terhadap kesalahan prediksi ekstrem (*Outliers*), sehingga memberikan gambaran stabilitas model yang lebih ketat (Chicco *et al.,* 2021). Sementara itu, MAPE ditambahkan untuk mengukur rata-rata kesalahan prediksi dalam bentuk persentase, sehingga memudahkan interpretasi performa model pada berbagai *event* dengan durasi yang berbeda-beda. Kombinasi keempat metrik ini memberikan gambaran yang komprehensif mengenai akurasi dan reliabilitas model prediksi.

Secara matematis, untuk data berukuran n, dengan nilai aktual y_i dan prediksi $\hat{y}_i$, metrik dapat dituliskan sebagai persamaan 2.1 sampai 2.4.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.3}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2.4}$$

dengan $\bar{y}$ adalah rata-rata dari nilai aktual y_i.

Untuk menjamin generalisasi model dan mengurangi risiko *overfitting*, evaluasi dilakukan menggunakan *k-Fold Cross-Validation*, yaitu membagi data menjadi k subset (*fold*) sehingga setiap subset pernah menjadi data uji dan sisanya menjadi data latih. Jika $Metric_j$ adalah nilai metrik pada *fold* ke-j, maka rata-rata metrik dapat dirumuskan dalam persamaan 2.5.

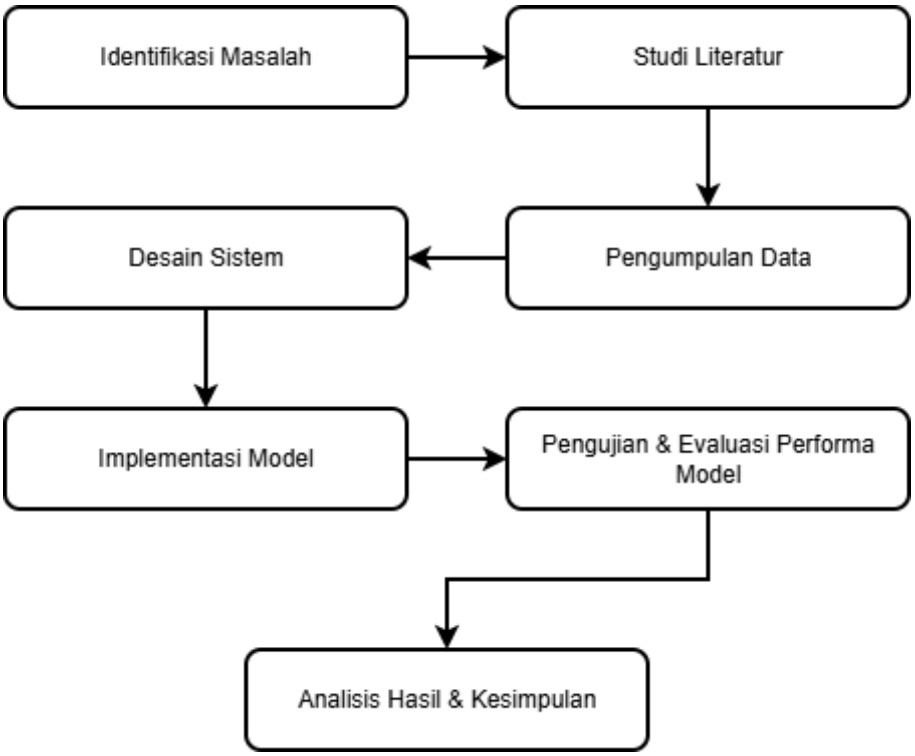
$$\text{Metric} = \frac{1}{k}\sum_{j=1}^{k}\text{Metric}_j \tag{2.5}$$

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Desain penelitian dirancang untuk memberikan kerangka kerja yang jelas dalam menjalankan sebuah penelitian, mulai dari identifikasi masalah hingga hasil akhir dan kesimpulan. Dengan adanya desain penelitian yang terstruktur, maka langkah-langkah penelitian dapat diatur secara sistematis, sehingga hasil yang diperoleh menjadi lebih akurat dan relevan dengan tujuan penelitian.



Gambar 3.1 Desain Penelitian

Desain penelitian pada Gambar 3.1, menggambarkan tahapan kerja penelitian yang disusun secara berurutan dari awal hingga akhir. Proses dimulai dari Identifikasi Masalah, yaitu penentuan fokus penelitian dan alasan mengapa prediksi

waktu tempuh pada *event Trail Running* perlu dilakukan. Setelah itu dilanjutkan Studi Literatur untuk menguatkan dasar teori, memperjelas variabel yang relevan, serta menentukan pendekatan penelitian yang sesuai. Berdasarkan dua tahap awal tersebut, penelitian masuk ke Desain Sistem, yaitu penyusunan kerangka kerja penelitian secara menyeluruh agar setiap langkah yang dilakukan konsisten dan dapat dipertanggungjawabkan.

Tahap berikutnya adalah Pengumpulan Data, yaitu mengumpulkan data yang dibutuhkan sebagai bahan penelitian, baik data utama perlombaan maupun data pendukung yang relevan. Setelah data tersedia, penelitian dilanjutkan ke Implementasi Model, yaitu penerapan metode yang dipilih untuk membangun sistem prediksi sesuai rancangan yang telah dibuat. Selanjutnya dilakukan Pengujian dan Evaluasi Performa Model untuk menilai apakah hasil prediksi sudah memenuhi tujuan penelitian dan untuk melihat tingkat ketepatan model berdasarkan ukuran evaluasi yang digunakan.

Tahap akhir adalah analisis Hasil & Kesimpulan, yaitu menginterpretasikan hasil pengujian, menjawab rumusan masalah, serta menyusun kesimpulan dan saran berdasarkan temuan penelitian. Dengan demikian, gambar tersebut menunjukkan alur yang runtut, sehingga setiap tahap menjadi dasar bagi tahap berikutnya

3.2 Sumber Data

Penelitian ini menggunakan data sekunder berupa dataset *UTMB World Race Data Sheet* periode 2014–2024 yang bersumber dari repositori Kaggle dalam format CSV. Penggunaan data sekunder ini dipilih karena mencakup rekam jejak historis yang luas, yang memungkinkan analisis mendalam pada tingkat *race-level*

(unit observasi berupa *event*). Fokus analisis ini bertujuan untuk mengisolasi pengaruh karakteristik lintasan dan lingkungan terhadap rata-rata waktu tempuh, sehingga faktor variansi performa individu atlet dapat diminimalisir (Ardigò *et al.,* 2020)

Secara keseluruhan, dataset awal terdiri dari 38.460 baris dengan 22 fitur yang mencakup dimensi identitas, temporal, fisik lintasan, partisipasi, hingga geospasial. Struktur data yang heterogen, terdiri dari 17 kolom numerik dan 5 kolom kategorikal menjadikan dataset ini sangat ideal untuk pendekatan *supervised regression* menggunakan model berbasis pohon Keputusan (Géron, 2022). Detail mengenai kategorisasi fitur tersebut dirangkum dalam tabel 3.1.

Tabel 3.1 *Feature* Dataset

Kategori	**Nama Fitur**	**Tipe Data**	**Deskripsi**
Identitas	race_uid, race_title, raw_location	Object	ID unik, nama perlombaan, dan lokasi tekstual.
Lintasan	distance, elevation_gain, elevation_*variance*	Float	Jarak total (km), total tanjakan (m), dan variansi elevasi.
Geografis	latitude, longitude, *country*, *Continent*	Float/Obj	Koordinat GPS, lokasi negara, dan benua.
Temporal	year, day	Float	Tahun dan hari ke-n dalam setahun (1-366).
Partisipasi	n_participants, n_women, n_countries	Integer	Jumlah total pelari, pelari wanita, dan asal negara.
Target	mean_finish_time	Float	Rata-rata waktu finish (dalam jam).

Dalam eksplorasi awal, ditemukan *Missing value* pada beberapa kolom (misalnya *Continent*) dan missing minimal pada target mean_finish_time. Namun, penelitian ini tidak menggunakan fitur lokasi berbasis nama (*country*/*Continent*) sebagai fitur model. Variasi antar wilayah direpresentasikan melalui karakteristik lintasan dan variabel lingkungan (cuaca) hasil integrasi spasio-temporal, bukan

label geografis (Kuhn & Johnson, 2013). Sebagai gambaran visual, Tabel 3.2 menyajikan sampel acak dari dataset tersebut sebelum dilakukan proses integrasi lebih lanjut.

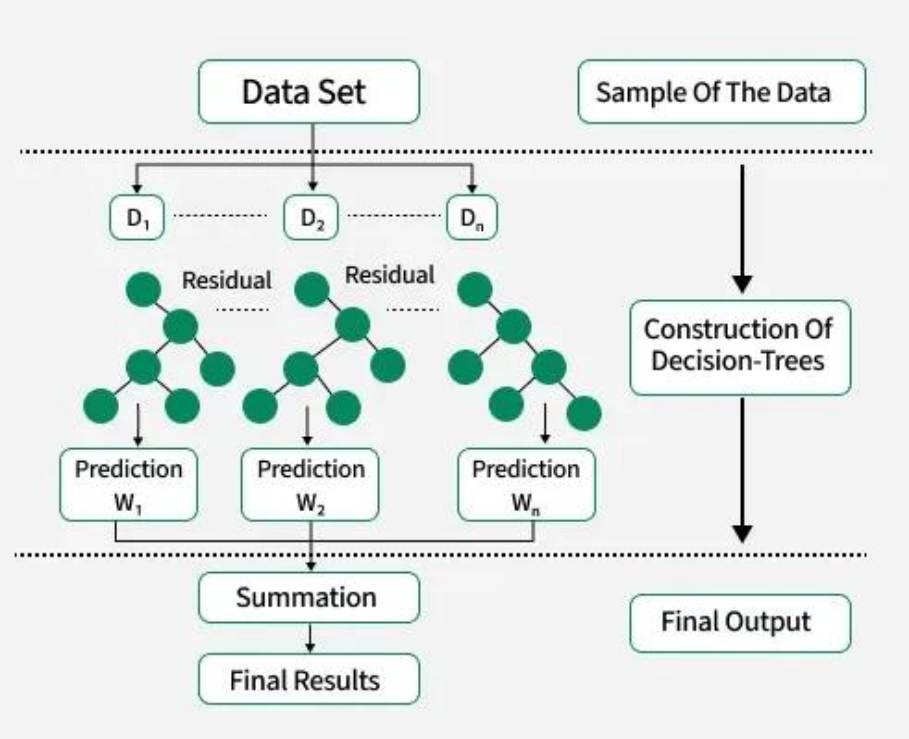
Tabel 3.2 Contoh Isi Dataset

Race Title	**Distance**	**Elev Gain**	***Continent***	**Year**	**Day**	***Mean finish time***
UTMB®	171.5	10040	Europe	2023	244	32.45
CCC®	101.3	6100	Europe	2022	238	19.12
Lavaredo Ultra Trail	120	5800	Europe	2021	176	22.3
Tarawera 102k	102.4	3089	Oceania	2019	40	14.55
Western States 100	161	5500	N. America	2018	174	26.1
Doi Inthanon 100	168	9800	Asia	2023	342	35.2
Eiger Ultra Trail	101	6700	Europe	2017	196	21.05
Mozart 100	105	5400	Europe	2022	169	16.8
Transgrancanaria	126	6820	Europe	2020	65	24.15
Ultra-Trail Mt. Fuji	165	7942	Asia	2015	268	31.9

Kombinasi antara keragaman fitur geografis dan metrik fisik lintasan di atas memungkinkan model *Machine learning* untuk menangkap variansi data lintas benua secara lebih akurat (Géron, 2022; Kuhn & Johnson, 2013).

3.3 Arsitektur Extreme Gradient Boosting (XGBoost)

XGBoost merupakan pengembangan dari *Gradient Boosted Decision Trees* yang dirancang untuk performa tinggi pada data tabular serta mendukung regularisasi untuk menekan *overfitting* (Chen & Guestrin, 2016).



Gambar 3.2 Arsitektur XGBoost

Seperti pada Gambar 3.2 model memprediksi output sebagai penjumlahan dari sejumlah pohon keputusan seperti pada persamaan 3.1.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), f_k \in \mathcal{F} \tag{3.1}$$

Keterangan:
$\mathbf{x}_i$= vektor fitur untuk data ke-i
$\hat{y}_i$= nilai prediksi
K= jumlah pohon
f_k= fungsi pohon ke-k yang memetakan fitur ke nilai daun

3.3.1 Fungsi objektif dan regularisasi

XGBoost meminimalkan fungsi objektif ter-regularisasi

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3.2}$$

dengan komponen regularisasi:

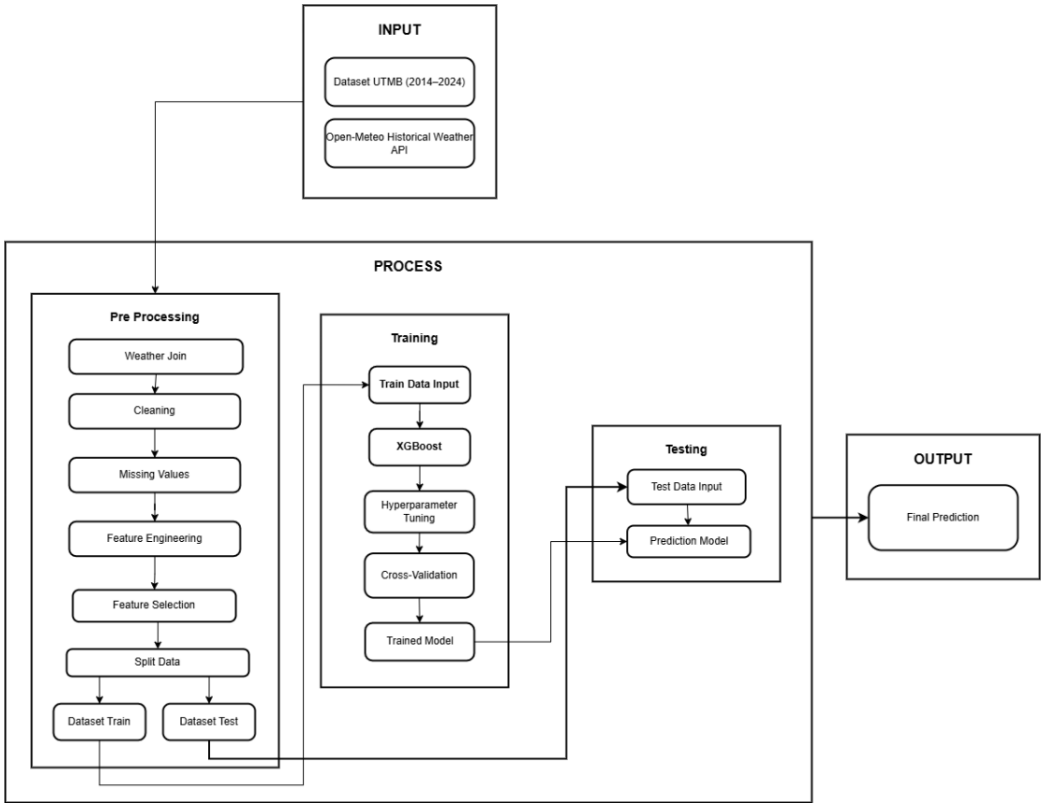
$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3.3}$$

Keterangan:
n= jumlah data latih
$l(y_i, \hat{y}_i)$= fungsi loss (untuk regresi umumnya kuadrat *error*)
$\Omega(f)$= penalti kompleksitas pohon
T= jumlah daun pada pohon
w_j= bobot pada daun ke-j
γdan λ= parameter regularisasi

XGBoost mengoptimalkan fungsi objektif menggunakan aproksimasi Taylor orde-2 sehingga pembentukan pohon lebih stabil dan efisien (Chen & Guestrin, 2016).

3.4 Desain Sistem

Desain sistem dalam penelitian ini dirancang untuk memodelkan prediksi performa pelari pada ajang *Ultra-Trail du Mont-Blanc* (UTMB) menggunakan pendekatan *Machine learning* berbasis algoritma XGBoost. Alur sistem dibagi menjadi tiga tahapan utama: *Input, Process,* dan *Output* seperti pada Gambar 3.3.



Gambar 3.3 Desain Sistem

1. *Input* Data

a. Sistem mengintegrasikan dua sumber data utama untuk membangun fitur yang komprehensif:

b. Dataset UTMB (2014–2024): Merupakan data historis perlombaan yang mencakup profil lintasan, kategori lomba, dan variabel pelari yang tersedia dalam dataset.

c. *Open-Meteo Historical Weather API*: Digunakan untuk mengekstraksi data cuaca historis berdasarkan koordinat lokasi dan waktu spesifik saat perlombaan berlangsung, guna memperkaya variabel lingkungan dalam model.

2. *Process* (Tahapan Pemrosesan)

Bagian proses merupakan inti dari sistem yang terdiri dari tiga modul utama yang saling berkaitan:

a. *Pre-Processing* (Pra-pemrosesan Data)

Tahap ini bertujuan untuk mentransformasi data mentah menjadi format yang optimal bagi model XGBoost dengan urutan sebagai berikut:

- *Weather join*: Melakukan penggabungan (*merging*) data cuaca ke dalam dataset utama UTMB berdasarkan kunci waktu dan lokasi *event*.
- *Cleaning*: Membersihkan dataset dari data duplikat, inkonsistensi unit, serta menangani *Outlier* yang tidak valid.
- *Missing values*: Menangani nilai kosong melalui teknik imputasi atau eliminasi sesuai dengan aturan yang telah ditetapkan dalam penelitian.
- *Feature Engineering*: Membentuk fitur-fitur baru (seperti rasio elevasi atau fitur musiman) untuk meningkatkan kemampuan prediksi model.
- *Feature Selection & Scenario Mapping*: Pada tahap ini, fitur dipilih untuk menghindari *data leakage*. Logika skenario penelitian (misalnya Skenario A, B, C, D) diterapkan di sini untuk menentukan himpunan fitur yang akan diuji.
- *Split* Data: Pembagian dataset menjadi *Dataset Train* untuk pembelajaran dan *Dataset Test* untuk pengujian independen.

b. *Training (*Pelatihan Model)

Proses pelatihan dilakukan menggunakan *Dataset Train* dengan tahapan:

- XGBoost Implementation: Menginisialisasi algoritma *XGBoost Regressor*.
- *Hyperparameter Tuning*: Mengoptimalkan parameter model (seperti *learning rate*, *max depth*, dan *n_estimators*) menggunakan teknik seperti *RandomizedSearchCV*.
- *Cross-Validation*: Memastikan model memiliki generalisasi yang baik dan menghindari *overfitting*.
- Trained Model: Menghasilkan model final yang telah teroptimasi untuk setiap skenario penelitian.

c. *Testing* (Pengujian Model)

Model yang telah dilatih kemudian dievaluasi menggunakan *Dataset Test*:

- Test Data *Input*: Memasukkan data uji yang belum pernah dilihat oleh model.
- *Prediction* Model: Model melakukan inferensi untuk memprediksi nilai target berdasarkan fitur-fitur pada data uji.

3. *Output*

Keluaran utama dari sistem ini adalah Final Prediction, yaitu nilai *mean finish time* pada level *event*. Hasil prediksi ini kemudian menjadi dasar untuk perhitungan metrik evaluasi seperti *MAE (Mean Absolute Error)* dan *RMSE (Root Mean Square Error)*, serta analisis interpretabilitas menggunakan SHAP untuk memahami pengaruh setiap fitur terhadap performa pelari.

3.5 *Preprocessing* Data

Tahapan *preprocessing* data bertujuan untuk menyiapkan dataset *event Trail Running* agar siap digunakan dalam proses pelatihan model *Machine learning*. *Dataset* yang digunakan merupakan data sekunder pada tingkat *event* (*race-level*), sehingga setiap baris data merepresentasikan satu *event Trail Running*. *Preprocessing* dilakukan untuk memastikan data bersih, konsisten, dan sesuai dengan kebutuhan algoritma *tree-based models*.

3.5.1 Definisi Variabel Target dan Transformasi Log

Target prediksi adalah *mean_finish_time* dalam satuan jam. Untuk menstabilkan varians dan mengurangi pengaruh *Outlier*, target dapat ditransformasikan menggunakan fungsi log. Seluruh metrik evaluasi dihitung pada skala waktu asli (jam) setelah dilakukan proses *inverse-transform* pada hasil prediksi model. Dengan demikian, interpretasi kesalahan prediksi tetap konsisten dalam satuan jam. Dalam penelitian ini, target dipertimbangkan menggunakan transformasi menggunakan persamaan 3.4.

$$y^* = log(1 + y) \tag{3.4}$$

Transformasi $\log(1 + y)$ dipilih agar aman untuk nilai kecil dan menjaga kestabilan numerik. Prediksi model kemudian dikembalikan ke skala jam dengan invers menggunakan persamaan 3.5.

$$\hat{y} = \exp(\hat{y}^*) - 1 \tag{3.5}$$

Sebagai catatan, XGBoost juga menyediakan objektif reg:squaredlog*error* yang mengoptimalkan *error* pada ruang log, namun pendekatan penelitian ini

menekankan transformasi target yang eksplisit agar interpretasi *pipeline* lebih jelas dalam proposal.

3.5.2 Integrasi Data Cuaca Spasio Temporal

Dataset UTMB tidak menyediakan informasi cuaca pada saat perlombaan. Oleh karena itu, penelitian menambahkan variabel cuaca menggunakan *Open-Meteo Historical Weather API*, yaitu layanan cuaca historis berbasis dataset *reanalysis (Open-Meteo, 2026)*. Integrasi dilakukan secara spasio-temporal: cuaca ditentukan oleh lokasi (koordinat) dan waktu pelaksanaan *event*. Koordinat *latitude* dan *longitude* digunakan semata-mata sebagai kunci penarikan cuaca, kemudian dihapus setelah proses *Weather join* agar tidak menjadi fitur model. Pengambilan dilakukan pada skala harian dengan:

1. *start_date = event_date,*
2. *end_date = event_date*,

sehingga setiap *event* memperoleh satu set variabel cuaca pada hari perlombaan (Open-Meteo, 2026). Secara empiris, kondisi cuaca seperti presipitasi dapat memengaruhi waktu finis pada lomba lintas alam sehingga integrasi cuaca relevan untuk pemodelan durasi *event* (Wilson & Wilson, 2024).

3.5.3 Pembersihan Data dan Penanganan *Missing value*s

Pembersihan data dilakukan untuk menjaga validitas *input* model, meliputi:

1. Menghapus baris dengan target kosong (*mean_finish_time* missing) karena tidak dapat digunakan dalam pelatihan regresi.

2. Menyaring nilai tidak logis (misalnya *distance* > 0, *elevation_gain* ≥ 0, *day* 1–366, serta rentang koordinat valid).

Karena fitur lokasi berbasis nama (*country* dan *Continent*) tidak digunakan dalam model, tidak dilakukan pemetaan "*Unknown*" untuk kolom tersebut. Penanganan missing difokuskan pada prediktor numerik (terutama cuaca). Nilai missing dipertahankan sebagai NaN karena XGBoost bersifat *sparsity-aware* dan mampu menangani missing secara internal melalui arah default pada *split* (Chen & Guestrin, 2016). Bila dilakukan imputasi (opsional *Robustness*), prosesnya mengikuti prinsip *Fit* pada data latih dan terapkan pada data uji untuk mencegah *data leakage* (Kuhn & Johnson, 2013).

3.5.4 *Feature Engineering* Berbasis ITRA dan *Gradien Lintasan*

Feature engineering difokuskan pada representasi beban lintasan sesuai konsep *ITRA (International Trail Running Association)*.

1. KM *Effort*

Feature engineering difokuskan pada representasi beban lintasan sesuai konsep *ITRA (International Trail Running Association)*. *ITRA (International Trail Running Association)* mendefinisikan *km-effort* sebagai kombinasi jarak dan elevasi positif, sehingga 1 km jarak setara 1 km-*effort* dan setiap 100 m elevasi positif menambah 1 km-*effort* (International *Trail Running* Association, 2022). Seperti pada persamaan 3.6.

$$KM_{effort} = \text{distance} + \left(\frac{\text{EG}}{100}\right) \tag{3.6}$$

2. EG/km (Elevation Gain per Kilometer)

$$EG/km = \frac{\text{EG}}{\text{distance}} \tag{3.7}$$

EG/km merupakan indikator "intensitas tanjakan" per km. Dalam praktik interpretasi, Elevasi merupakan prediktor penting performa *ultra-endurance* sehingga fitur berbasis elevasi relevan untuk prediksi durasi (Knechtle *et al.,* 2025). EG/km juga sering diasosiasikan dengan kemiringan rata-rata (*gradien*). Jika dikonversi ke persen kemiringan rata-rata:

$$\text{Gradien}_{\text{avg}}(\%) = \left(\frac{EG/km}{10}\right) \tag{3.8}$$

Catatan: ini merupakan proksi (rata-rata global), karena lintasan trail memiliki kombinasi tanjakan dan turunan yang tidak seragam.

3.5.5 Seleksi Fitur

Seleksi fitur dilakukan bertahap untuk memastikan model hanya menggunakan informasi yang tersedia sebelum perlombaan dimulai dan menghindari *data leakage* (Kuhn & Johnson, 2013). Tahapan seleksi terdiri dari (1) eliminasi awal dan (2) perangkingan fitur untuk kebutuhan skenario efisiensi.

1. Tahap Eliminasi Awal.

Fitur identitas *event* serta variabel pasca-lomba dieliminasi. Selain itu, fitur identitas lokasi berbasis nama (*country*/*Continent*) tidak digunakan agar model lebih general dan tidak "menghafal" lokasi. Variabel pasca lomba seperti *winning_time, Last Time,* dan N DNF dieliminasi sepenuhnya karena hanya diketahui setelah *event* selesai dan berpotensi menjadi kebocoran jawaban. Selain itu, *event_date* tidak digunakan sebagai fitur karena digantikan oleh fitur musiman

seperti *month* dan *day_of_year*. Setelah proses integrasi cuaca selesai, koordinat *latitude* dan *longitude* dihapus untuk mengurangi redundansi. Pada tahap ini, *Continent* juga dieliminasi untuk menghindari redundansi dan menjaga fitur lokasi tetap sederhana serta relevan untuk skenario *end user*. Aturan eliminasi dijelaskan pada Tabel 3.3.

Tabel 3.3 Aturan Eliminasi Fitur

Kategori	**Nama Fitur**	**Status**	**Alasan Eliminasi**
Identitas	*race_uid, race_title, raw_location*	Drop	ID atau teks unik; tidak memiliki nilai prediktif general; risiko *overfitting.*
Leakage	*winning_time, Last Time*, N DNF	Drop	Informasi pasca-lomba (bocoran jawaban); tidak tersedia saat melakukan prediksi sebelum lomba dimulai.
Redundan	*latitude, longitude*	Drop	Hanya digunakan untuk *Weather join*.
Diluar *Scope*	*Continent*,*country*	Drop	Fokus penelitian di karakteristik lintasan dan cuaca, jadi lokasi tidak digunakan karena bias.
Temporal	*event*_date.	Drop	Digantikan oleh fitur hasil ekstraksi seperti month.
Target	*mean_finish_time*	Target	Variabel dependen utama yang akan diprediksi oleh model.

2. Fitur yang Dipertahankan

Setelah eliminasi awal, tersisa himpunan fitur valid yang merupakan gabungan fitur fisik lintasan, statistik peserta, fitur temporal, fitur turunan

(*engineered)*, serta fitur cuaca. Daftar fitur yang dipertahankan ditunjukkan pada Tabel 3.4.

Tabel 3.4 Daftar Fitur yang Dipertahankan

Kelompok	Nama Fitur (Contoh/Utama)	Status	Keterangan Singkat
Temporal	*Month, year, day_of_year.*	Keep	Menangkap tren pola musiman (misal: suhu udara).
Lintasan (Fisik)	*distance_km, elevation_gain_m, elevation_variance, location_elevation_m, admin_level.*	Keep	Determinan utama beban lintasan dan durasi lari.
Partisipasi	*n_participants, n_women, n_countries.*	Keep	Proksi kepadatan kompetisi dan level keragaman *event*.
Kategorikal	*race_category.*	Keep	Kelas atau tipe lomba (perlu di-encoding ke angka).
Engineered	km_*effort*, eg_per_km.	Keep	Ringkasan kompleksitas: gabungan jarak + elevasi dan intensitas tanjakan.
Cuaca	*wx_temp_max, wx_temp_min, wx_precip_sum, wx_wind_speed, wx_temp_mean.*	Keep	Variabel lingkungan dinamis yang mungkin mempengaruhi performa fisik.

3. Tahap Perengkingan Fitur (*Feature importance Ranking)*

Setelah diperoleh *Full Valid Feature Set*, fitur-fitur valid diurutkan berdasarkan tingkat kepentingannya dengan melatih Model A menggunakan *hyperparameter* terbaik hasil proses optimasi. Nilai kepentingan fitur diekstrak menggunakan metrik *Gain Importance*, yaitu ukuran kontribusi rata-rata suatu fitur dalam meningkatkan kualitas *split* pada pohon keputusan.

Hasil perankingan tersebut digunakan sebagai dasar pembentukan subset fitur pada Model C, yaitu dengan memilih sejumlah fitur teratas (misalnya Top-10 dari total fitur yang tersedia) untuk menguji efisiensi model terhadap reduksi dimensi *input*. Selain itu, hasil ranking juga menjadi pertimbangan dalam penyusunan komposisi fitur sederhana pada Model D, yang dirancang untuk

mensimulasikan skenario implementasi dengan *input* minimal dan mudah diakses oleh pengguna akhir.

3.5.6 Pembagian Data dan Posisi Train–Test *Split*

Pembagian data dilakukan untuk memastikan evaluasi performa model bersifat independen, meminimalkan bias estimasi, serta menjaga integritas *pipeline* dari potensi *data leakage*. Strategi pembagian data dalam penelitian ini mengikuti praktik standar pengembangan model prediktif sehingga hasil dapat direplikasi dan dibandingkan secara adil antar skenario.

1. Rasio Data

Penelitian mengevaluasi tiga skenario *train-test split* yaitu 90:10, 80:20, dan 70:30.

2. Reprodusibilitas (*Random state* Konstan)

Pembagian data menggunakan nilai *Random state* yang tetap untuk memastikan hasil pembagian konsisten pada setiap eksperimen, sehingga seluruh skenario pengujian dapat direplikasi dan hasilnya dapat dibandingkan secara objektif.

3. *Stratified* y-bins untuk Menjaga Distribusi Target

Untuk menjaga distribusi target pada regresi, pembagian data dilakukan secara *stratified* dengan membentuk *y-strata (bins)* menggunakan kuantil pada nilai target di data latih. Secara operasional, setelah data dibagi menjadi *train* dan *test*, nilai target pada data latih (*y_train*) dikelompokkan ke dalam K bin kuantil (misalnya K=10), lalu label bin tersebut digunakan untuk menjaga keseimbangan distribusi target pada proses validasi silang (*Cross-Validation*) dan/atau pembagian

yang memerlukan stratifikasi. Pembuatan bin dilakukan berdasarkan data latih agar tidak memasukkan informasi dari data uji.

4. Pencegahan Data Leakage dan Posisi *Split*.

Dalam *pipeline* penelitian, *train–test split* ditempatkan setelah pembersihan data awal dan *feature engineering* yang bersifat deterministik (misalnya perhitungan km_*effort* dan eg_per_km). Namun, *split* dilakukan sebelum transformasi yang memerlukan proses "*fit*" pada data, seperti imputasi statistik (bila digunakan) dan pembentukan struktur kolom hasil encoding. Prinsip ini penting agar parameter transformasi hanya dipelajari dari data latih dan kemudian diterapkan ke data uji (*fit on train, apply on test*), sehingga informasi dari data uji tidak "bocor" ke dalam proses pelatihan model, sejalan dengan praktik terbaik dalam pengembangan model *machine learning* (Kuhn & Johnson, 2013).

3.6 *Training* Model Dan Optimasi *Hyperparameter*

Model utama dalam penelitian ini dibangun menggunakan algoritma XGBRegressor, yaitu implementasi *gradient boosted decision trees* yang dioptimalkan untuk efisiensi komputasi dan performa pada tugas regresi. Sejalan dengan desain sistem pada Gambar 3.2, proses pelatihan diawali dengan menjalankan Model A, yaitu model yang menggunakan seluruh fitur valid yang tersedia sebelum perlombaan berlangsung. Model A diposisikan sebagai *Baseline* atau *benchmark* untuk membandingkan performa model lainnya karena merepresentasikan batas atas pemanfaatan informasi fitur yang paling lengkap.

Untuk menjamin bahwa perbandingan antar model bersifat adil dan tidak dipengaruhi oleh perbedaan konfigurasi algoritma, penelitian ini melakukan

optimasi *hyperparameter* terlebih dahulu pada Model A. Kombinasi *hyperparameter* terbaik yang diperoleh dari proses tersebut kemudian digunakan sebagai parameter tetap (*fixed*) pada pelatihan dan evaluasi Model B, Model C, dan Model D. Dengan pendekatan ini, perubahan performa antar model dapat diatribusikan secara spesifik pada perbedaan komposisi fitur, bukan pada variasi hasil *Tuning*.

Selain berfungsi sebagai *Baseline*, Model A juga digunakan untuk menghitung *Feature importance* berbasis *gain*, yang kemudian menjadi dasar dalam pembentukan subset fitur pada Model C serta penyusunan komposisi fitur sederhana pada Model D.

3.6.1 Spesifikasi Parameter XGBRegressor

Parameter XGBRegressor dikelompokkan menjadi parameter pembelajaran, kompleksitas pohon, *subsampling*, regularisasi, dan konfigurasi umum. Ringkasan parameter disajikan pada Tabel 3.5.

Tabel 3.5 Spesifikasi Parameter XGBRegressor

Kelompok	Parameter	Fungsi/Peran	Catatan Penggunaan
Pembelajaran	*learning*_rate	Mengatur besar langkah pembaruan tiap iterasi	Nilai kecil (0,01–0,2) lebih stabil.
	n_estimators	Jumlah total pohon (boosting rounds)	Dikontrol via CV/early stopping (300–2000).
Kompleksitas	max_depth	Kedalaman maksimum pohon	Depth besar (3–10) menangkap pola kompleks.
	min_child_weight	Ambang minimum bobot node untuk *split*	Nilai besar (1–10) membuat model konservatif.
	gamma	Penurunan loss minimum untuk *split*	Nilai besar (0–1) membuat *split* lebih selektif.
Sampling	subsample	Sampling baris untuk tiap pohon	Rasio 0,6–1,0 untuk mengurangi noise.
	colsample_bytree	Sampling fitur untuk tiap pohon	Rasio 0,6–1,0 untuk menekan varians.
	colsample_bylevel	Sampling fitur per level pohon	Tambahan regularisasi (opsional).

Kelompok	Parameter	Fungsi/Peran	Catatan Penggunaan
Regularisasi	reg_alpha (L1)	Menekan bobot fitur kurang penting	Digunakan untuk menciptakan sparsity.
	reg_lambda (L2)	Menstabilkan bobot model	Mencegah *overfitting* ekstrem (≥1).
Konfigurasi	objective	Fungsi objektif regresi	Menggunakan reg:squared*error*.
	random_state	Seed konsistensi hasil	Ditentukan angka konstan (misal: 42).

3.6.2 Optimasi *Hyperparameter* pada Model A

Optimasi dilakukan menggunakan *RandomizedSearchCV* pada data latih dengan *5-fold Cross-Validation* (Pedregosa *et al.,* 2011) Scoring metric ditetapkan *MAPE (neg_mean_absolute_percentage_error)* agar objective *Tuning* selaras dengan metrik utama evaluasi dalam satuan jam. MAE tetap dihitung dan dilaporkan sebagai metrik pendukung persentase, namun bukan tujuan utama optimasi (Chai & Draxler, 2014).

Tabel 3.6 Ruang *Hyperparameter* untuk *RandomizedSearchCV*

Parameter	Tipe	Kandidat / Rentang yang Diuji
*learning*_rate	Diskrit	{0,01; 0,05; 0,1}
n_estimators	Diskrit	{500;1000; 1500}
max_depth	Diskrit	{4 , 6, 8, 10}
min_child_weight	Diskrit	{1, 3, 5, 7, 10}
gamma	Diskrit	{0; 0,1; 0,2; 0,5}
subsample	Diskrit	{0,6; 0,7; 0,8; 0,9; 1,0}
colsample_bytree	Diskrit	{0,6; 0,7; 0,8; 0,9; 1,0}
reg_alpha	Diskrit	{0; 1e-4; 1e-3; 1e-2; 0,1; 1,0}
reg_lambda	Diskrit	{0,5; 1,0; 2,0; 5,0; 10,0}

Untuk tabel konfigurasi *RandomizedSearchCV* kurang lebih seperti Tabel 3.7.

Tabel 3.7 Konfigurasi *RandomizedSearchCV*

Komponen	Nilai / Setelan	Keterangan
Metode Pencarian	*RandomizedSearchCV*	Efisien untuk ruang parameter yang luas.
n_iter	10	Jumlah kombinasi parameter yang diuji.
Cross-Validation	5-fold CV	Validasi silang pada data latih.
Scoring Metric	neg_mean_absolute_percentage_*error*	Optimasi berdasarkan MAPE terendah.
Data Source	Train Set	CV dilakukan hanya pada data latih.
Paralelisasi	n_jobs	Disesuaikan dengan kapasitas CPU.

Karena Setelah kombinasi terbaik diperoleh, Model A dilatih ulang pada seluruh data latih dan dievaluasi pada data uji menggunakan metrik MAE, RMSE, R^2, dan MAPE.

3.6.3 Training Skenario

Setelah parameter terbaik diperoleh dari Model A, penelitian melanjutkan proses pelatihan pada model lainnya dengan menggunakan konfigurasi *hyperparameter* yang sama. Pendekatan ini dilakukan agar dampak perubahan komposisi fitur dapat diukur secara objektif, sehingga perbedaan performa yang muncul benar-benar mencerminkan variasi informasi *input*, bukan perbedaan hasil optimasi model.

1. Model B: Menghilangkan seluruh fitur cuaca (wx_*) untuk mengukur penurunan akurasi ketika variabel lingkungan dinamis tidak disertakan dalam proses prediksi.

2. Model C: Menggunakan sejumlah fitur teratas (misalnya 10–15 fitur) berdasarkan peringkat *gain importance* yang dihitung dari Model A, dengan tujuan menguji efisiensi model terhadap reduksi jumlah fitur tanpa kehilangan akurasi secara signifikan.
3. Model D: Menggunakan fitur sederhana yang mudah diketahui sebelum perlombaan, seperti jarak (*distance*), elevasi (*elevation_gain*), kategori lomba, dan bulan pelaksanaan, untuk menilai kelayakan implementasi model dalam skenario aplikasi *end-user* dengan *input* minimal.

3.7 Evaluasi Model

Kinerja model dievaluasi menggunakan MAE, RMSE, R^2, dan MAPE. MAE digunakan sebagai metrik utama karena interpretasinya langsung dalam satuan jam, sedangkan RMSE lebih sensitif terhadap *error* besar sehingga membantu menilai risiko *Outlier* (Chai & Draxler, 2014). R^2digunakan untuk menunjukkan proporsi variansi target yang dapat dijelaskan model (Chicco *et al.,* 2021). MAPE dilaporkan sebagai metrik pendukung dalam persen untuk perbandingan lintas *event* dengan durasi berbeda (Chai & Draxler, 2014).

Validasi *k-fold (k=5)* dilakukan pada data latih dan mengikuti *stratified y-bins* agar distribusi target konsisten antar *fold* (Kuhn & Johnson, 2013).

3.7.1 Metrik Evaluasi Regresi

1. MAE (*Mean Absolute Error*)

MAE mengukur rata-rata kesalahan absolut antara prediksi dan nilai aktual dalam satuan jam. Metrik ini mudah diinterpretasikan karena menunjukkan rata-rata selisih waktu asli yang dihasilkan model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.9}$$

Keterangan:
y_i: nilai aktual *mean finish time* pada data ke-i
$\hat{y}_i$: nilai *mean finish time* hasil prediksi model pada data ke-i
n: jumlah total data pada dataset uji
$|y_i - \hat{y}_i|$: selisih absolut antara nilai aktual dan nilai prediksi

2. RMSE (*Root Mean Square Error*)

RMSE memberikan penalti lebih besar pada kesalahan prediksi yang ekstrem. Nilai RMSE yang rendah mengindikasikan bahwa model jarang melakukan kesalahan prediksi yang fatal (sangat jauh dari nilai asli).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.10}$$

Keterangan:
y_i: nilai aktual *mean finish time* pada data ke-i
$\hat{y}_i$: nilai *mean finish time* hasil prediksi model pada data ke-i
n : jumlah total data pada dataset uji
$(y_i - \hat{y}_i)^2$: kuadrat selisih antara nilai aktual dan nilai prediksi
$\sqrt{...}$: akar kuadrat untuk mengembalikan satuan ke satuan waktu asli (jam)

3. R^2 Score (Koefisien Determinasi)

Metrik ini mengukur seberapa besar persentase variasi waktu finis yang dapat dijelaskan oleh fitur-fitur (jarak, elevasi, cuaca) yang digunakan dalam model.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3.11}$$

Keterangan:
y_i: nilai aktual *mean finish time*
$\hat{y}_i$: nilai prediksi model
$\bar{y}$: rata-rata nilai aktual *mean finish time*

4. MAPE (*Mean Absolute Percentage Error*)

MAPE mengukur rata-rata kesalahan prediksi dalam bentuk persentase sehingga lebih mudah dibandingkan lintas *event* dengan durasi yang berbeda.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3.12}$$

Keterangan:
y_i: nilai aktual *mean finish time* pada data ke-i
$\hat{y}_i$: nilai *mean finish time* hasil prediksi model pada data ke-i
n: jumlah total data pada dataset uji

3.7.2 Validasi Model dengan k-Fold Cross-*Validation*

Untuk menjamin validitas dan stabilitas hasil, penelitian ini menerapkan teknik k-*Fold Cross-Validation* dengan nilai $k = 5$. Proses ini bekerja dengan membagi data latih menjadi lima bagian (*folds*) yang sama besar. Pada setiap iterasi, empat bagian digunakan sebagai data latih dan satu bagian sisanya digunakan sebagai data validasi.

Proses ini diulang sebanyak lima kali sehingga setiap bagian data pernah diposisikan sebagai data validasi. Hasil akhir evaluasi diperoleh dari rata-rata nilai metrik di seluruh lipatan sesuai persamaan 3.13.

$$\bar{\text{Metric}} = \frac{1}{k}\sum_{i=1}^{k} \text{Metric}_i \tag{3.13}$$

Di mana Metric_i adalah nilai MAE, RMSE, atau R^2 pada lipatan ke-i . Penerapan metode ini sangat krusial untuk menghindari bias pemilihan data latih dan memastikan bahwa model XGBoost memiliki kemampuan generalisasi yang kuat saat menghadapi data perlombaan baru di luar dataset pelatihan.

3.8 Skenario Pengujian

Untuk menjawab tujuan penelitian secara komprehensif, pengujian dirancang dalam empat model eksperimen utama. Desain ini bertujuan untuk mengisolasi dan mengukur dampak keberadaan fitur cuaca serta efisiensi jumlah fitur terhadap akurasi prediksi *mean finish time*.

Dalam penelitian ini, Model A diposisikan sebagai *Baseline* atau *benchmark* karena menggunakan himpunan fitur valid yang paling lengkap sebelum perlombaan berlangsung. Optimasi *hyperparameter* dilakukan terlebih dahulu pada Model A menggunakan *RandomizedSearchCV* dengan 5-Fold *Cross-Validation* untuk memperoleh kombinasi parameter terbaik. Parameter terbaik tersebut kemudian digunakan sebagai konfigurasi tetap (*fixed*) pada Model B, Model C, dan Model D, sehingga perbandingan performa antar model merefleksikan perbedaan komposisi fitur, bukan perbedaan hasil proses *Tuning*.

Seluruh model dilatih menggunakan dataset latih yang sama dan dievaluasi pada porsi data uji yang identik (20% dari total populasi). Pendekatan ini diterapkan untuk menjamin bahwa perbandingan performa antar model bersifat adil, konsisten, dan objektif. Detail dari masing-masing model pengujian dirangkum dalam Tabel 3.8.

Tabel 3.8 Matriks Skenario Pengujian Model

Kode Uji	Komposisi Fitur	Deskripsi
Model A	*Full Feature Set*	Skenario utama dengan fitur lengkap untuk menentukan batas atas akurasi serta menghasilkan ranking gain importance.
Model B	Model A tanpa atribut cuaca (wx_*)	Mengukur penurunan akurasi ketika variabel lingkungan dinamis dihilangkan.
Model C	10 fitur teratas (berdasarkan ranking gain Model A)	Menguji performa model dengan subset fitur paling kontributif untuk efisiensi *input*/komputasi.
Model D	Fitur sederhana (dist, elev, cat, month)	Menilai kelayakan implementasi aplikasi end-user dengan *input* minimal (Jarak, Elevasi, Kategori, dan Bulan) tanpa bergantung pada data cuaca teknis atau identitas lokasi”

Melalui matriks pengujian ini, penelitian dapat membedah kontribusi masing-masing komponen secara mendalam. Hasil dari keempat skenario dianalisis menggunakan metrik MAE, RMSE, R^2, dan MAPE yang telah ditetapkan pada subbab evaluasi.

3.9 Analisis Interpretabilitas Model Menggunakan SHAP

Meskipun algoritma XGBoost memiliki akurasi yang tinggi, model ini sering kali dianggap sebagai *black-box* karena sulit dipahami bagaimana proses internalnya dalam menghasilkan prediksi (Lundberg *et al.,* 2020). Oleh karena itu, penelitian ini menggunakan metode *Shapley Additive Explanations* (SHAP) untuk

membedah kontribusi setiap variabel (seperti jarak, elevasi, dan cuaca) terhadap nilai prediksi *mean finish time*.

3.9.1 Konsep Dasar SHAP

SHAP didasarkan pada teori permainan (*game theory*) yang bertujuan untuk mendistribusikan kontribusi setiap fitur secara adil terhadap hasil akhir prediksi. Dalam konteks penelitian ini, SHAP membantu menjawab fitur mana yang paling dominan dalam mempercepat atau memperlambat durasi waktu finis sebuah perlombaan.

3.9.2 Rumus Atribusi SHAP

Inti dari metode SHAP adalah menjelaskan nilai prediksi sebagai jumlah aditif dari kontribusi setiap fitur *input*nya (Anastasiadou *et al.,* 2025). Secara sederhana, rumus untuk menjelaskan prediksi pada satu data tunggal adalah seperti pada persamaan 3.14.

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z_j' \tag{3.14}$$

Keterangan:
$g(z')$: Model penjelasan (*explanation model*) yang mewakili prediksi.
ϕ_0 : Nilai dasar (*base value*), yaitu rata-rata prediksi model pada seluruh dataset.
ϕ_j : Nilai SHAP untuk fitur ke-j, yang menunjukkan besarnya kontribusi fitur tersebut.
M : Jumlah total fitur yang digunakan dalam model.

3.9.3 Implementasi dan Visualisasi

Analisis dilakukan menggunakan *TreeExplainer*, sebuah pendekatan yang dioptimalkan untuk model berbasis pohon seperti XGBoost. Hasil analisis akan disajikan seperti berikut:

1. SHAP *Summary Plot*: Untuk melihat peringkat fitur yang paling berpengaruh secara global pada seluruh perlombaan *Trail Running*.
2. SHAP *Waterfall Plot*: Untuk menunjukkan bagaimana fitur-fitur pada satu *event* tertentu (secara lokal) bekerja sama sehingga menghasilkan prediksi waktu tempuh tersebut.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Bagian ini memaparkan hasil dari implementasi algoritma *Extreme Gradient Boosting* (XGBoost) untuk memprediksi *mean finish time* pada *event Trail Running* menggunakan dataset UTMB periode 2014–2024.

4.1.1 Hasil *Preprocessing* Data

Tahap *preprocessing* merupakan langkah krusial untuk menjamin kualitas *input* sebelum digunakan dalam pemodelan, mengingat kualitas data sangat menentukan performa akhir algoritma *Machine learning*.

```
[PILOT] Loading dataset: /content/utmb_with_weather_data.csv
[PILOT] Rows (raw): 38,398 | Cols: 31
[PILOT] Rows after pilot sampling (100%): 38,398
[PILOT] Rows after cleaning: 36,700
[PILOT] Feature engineering: created month, km_effort, eg_per_km
[PILOT] Anti-leakage filtering: dropped mandatory columns (if present)
[PILOT] Created y-strata bins: 10 bins
```

Gambar 4.1 Hasil *Preprocessing*

Gambar 4.1 menampilkan log sistem dari tahapan *pilot sampling* dan pembersihan data yang dilakukan secara otomatis melalui *pipeline* Python. Dari total 38.398 baris data mentah yang dimuat, sistem berhasil melakukan pembersihan (*cleaning*) sehingga menyisakan 36.700 baris data valid yang siap digunakan untuk pemodelan. Gambar tersebut juga mengonfirmasi keberhasilan proses *feature engineering* yang ditandai dengan terbentuknya kolom baru yaitu month, km_*effort*, dan eg_per_km. Selain itu, terlihat pula implementasi *anti-leakage filtering* untuk memastikan tidak ada fitur pasca-lomba yang masuk ke

dalam model, serta pembentukan 10 *y-strata bins* untuk kebutuhan *stratified split* agar distribusi target tetap terjaga seimbang.

Tabel 4.1 Highlight Data Setelah Feature Engineering

Race Title	**Distance (km)**	**Elev Gain (m)**	**KM *Effort***	**EG/km**	**Month**	***Mean finish time* (hours)**
UTMB®	171.5	10.040	271.90	58.54	8	32.45
CCC®	101.3	6.100	162.30	60.22	8	19.12
Lavaredo Ultra Trail	120.0	5.800	178.00	48.33	6	22.30
Western States 100	161.0	5.500	216.00	34.16	6	26.10

Sebelum ke bagian hasil training pada tabel 4.2 diperlihatkan pembagian fitur yang akan dilatih untuk setiap model.

Tabel 4.2 Pembagian Fitur Setiap Model

Kategori	**Nama Fitur**	**Model A**	**Model B**	**Model C**	**Model D**
Temporal	*year*	✓	✓	-	-
	day_of_year	✓	✓	-	-
	month	✓	✓	-	✓
Demografi Partisipan	*n_participants*	✓	✓	-	-
	n_women	✓	✓	✓	-
	n_countries	✓	✓	-	-
Fisik Lintasan & Lokasi	*race_category*	✓	✓	✓	✓
	distance_km	✓	✓	✓	✓
	elevation_gain_m	✓	✓	-	✓
	admin_level	✓	✓	✓	-
	location_elevation_m	✓	✓	✓	-
	elevation_variance	✓	✓	-	-
	eg_per_km	✓	✓	-	-

Kategori	Nama Fitur	Model A	Model B	Model C	Model D
	km_effort	✓	✓	✓	-
Kondisi Cuaca	*wx_temp_max*	✓	-	✓	-
	wx_temp_min	✓	-	✓	-
	wx_temp_mean	✓	-	✓	-
	wx_precip_sum	✓	-	✓	-
	wx_wind_speed	✓	-	-	-
Total	Jumlah Fitur	19 Fitur	14 Fitur	10 Fitur	4 Fitur

4.1.2 Hasil *Training* dan Optimasi *Hyperparameter*

Proses pelatihan model diawali dengan optimasi *hyperparameter* secara mendalam pada Model A (skenario fitur lengkap) menggunakan teknik *RandomizedSearchCV* dengan mekanisme *5-fold Stratified Cross-Validation* pada data latih. Langkah ini bertujuan untuk mencari titik keseimbangan optimal antara kompleksitas pohon dan kemampuan regularisasi guna mencegah *overfitting*.

Melalui pengujian terhadap 10 kombinasi parameter yang mencakup variabel pembelajaran, kompleksitas, hingga sampling, diperoleh nilai *negative Mean Absolute Percentage Error* (neg-MAPE) terbaik sebesar -0,0917. Nilai negatif ini merupakan representasi standar dari *Scikit-Learn* di mana skor yang lebih tinggi (mendekati nol) menunjukkan performa yang lebih baik.

Tabel 4.3 Parameter Terbaik dari Randomized Search CV

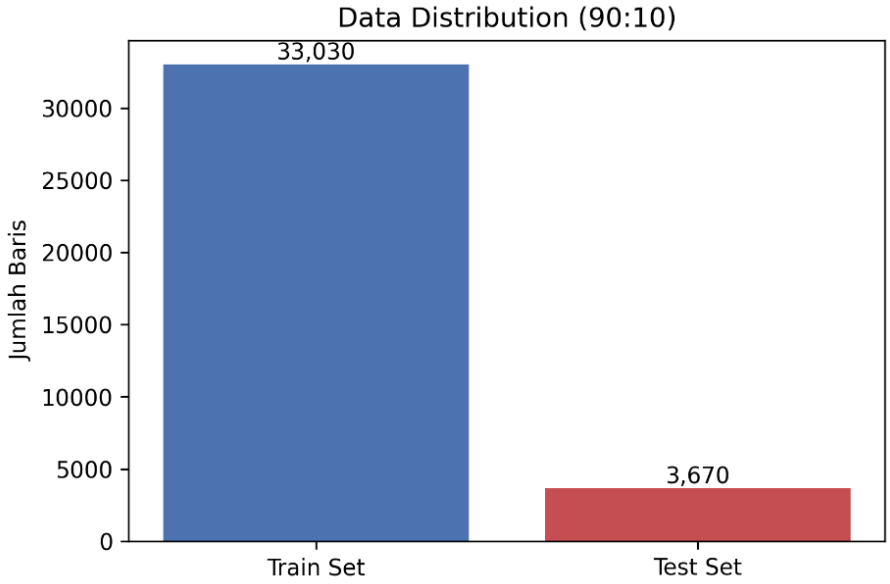
Kelompok	Parameter	Nilai Optimal
Pembelajaran	*learning_rate*	0.05
	n_estimators	1000
Kompleksitas	*max_depth*	8
	min_child_weight	3
	gamma	0.1
Sampling	*subsample*	0.8

Kelompok	Parameter	Nilai Optimal
	colsample_bytree	0.8
Regularisasi	*reg_alpha (L1)*	0.1
	reg_lambda (L2)	1

Parameter optimal tersebut kemudian dikunci (*fixed*) dan diterapkan secara konsisten pada seluruh skenario pelatihan model lainnya, mulai dari Model B hingga Model D. Strategi ini sangat krusial untuk memastikan bahwa setiap fluktuasi performa yang muncul di antara skenario benar-benar merefleksikan variasi komposisi fitur, bukan disebabkan oleh perbedaan hasil optimasi algoritma.

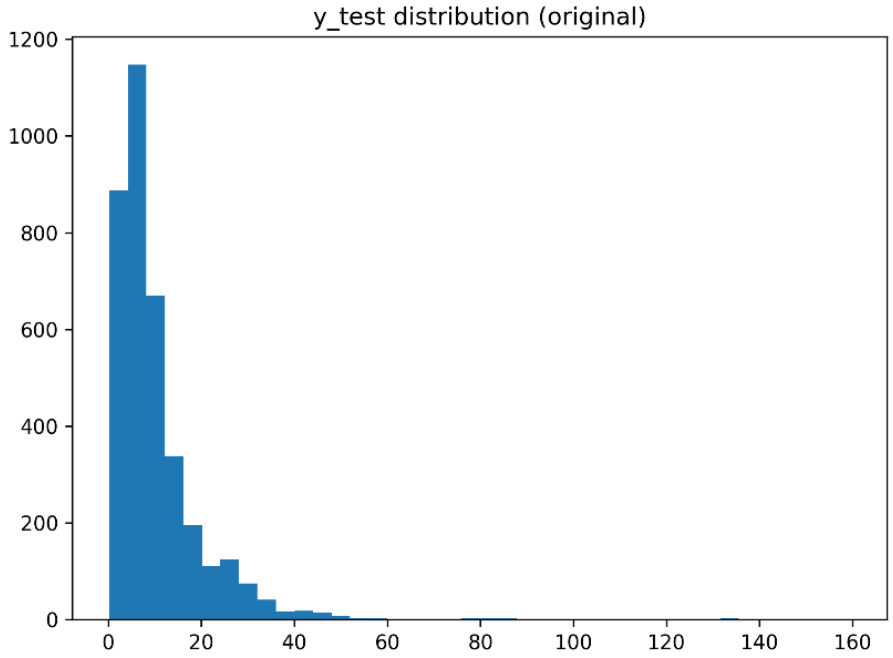
Dengan tercapainya nilai MAPE di kisaran 9%, model menunjukkan kapabilitas yang sangat andal dalam memprediksi durasi lomba bahkan sebelum dilakukan pengujian pada data independen.

4.1.3 Hasil Evaluasi *Split* Data 90 : 10



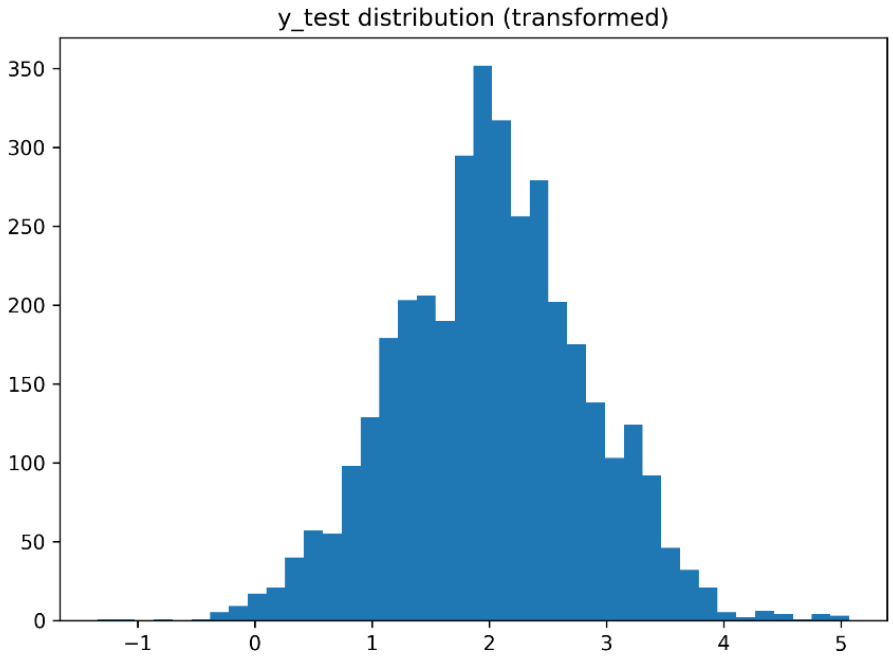
Gambar 4.2 Distribusi *Split* Data 90:10

Visualisasi pembagian data menunjukkan bahwa sebagian besar data digunakan untuk pelatihan model. Hal ini memberikan keuntungan berupa kemampuan model dalam menangkap pola data secara lebih komprehensif, yang berkontribusi terhadap performa prediksi yang lebih tinggi.



Gambar 4.3 Y test distribution 90:10 Original

Distribusi Grafik ini menunjukkan sebaran rata-rata waktu finis pada 10% data uji dari total populasi. Terlihat pola distribusi yang sangat condong ke kanan (*right-skewed*), sehingga konsentrasi data terbanyak berada pada rentang 5 hingga 20 jam. Adanya ekor panjang hingga melampaui 140 jam mengindikasikan keberadaan perlombaan kategori *ultra-trail* ekstrem yang jumlahnya jauh lebih sedikit dibandingkan lomba jarak menengah.



Gambar 4.4 Y *Test* Distribution 90:10 *Transformed*

Gambar 4.4 menampilkan hasil setelah data pada Gambar 4.2 dikenakan transformasi logaritmik ($log(1 + y)$). Terlihat bahwa distribusi data yang semula *skewed* kini berubah menjadi lebih simetris dan menyerupai distribusi normal (Gaussian). Transformasi ini krusial untuk membantu algoritma XGBoost dalam melakukan pemisahan (*split*) pada pohon keputusan secara lebih stabil dan mengurangi sensitivitas berlebih terhadap nilai-nilai ekstrem.

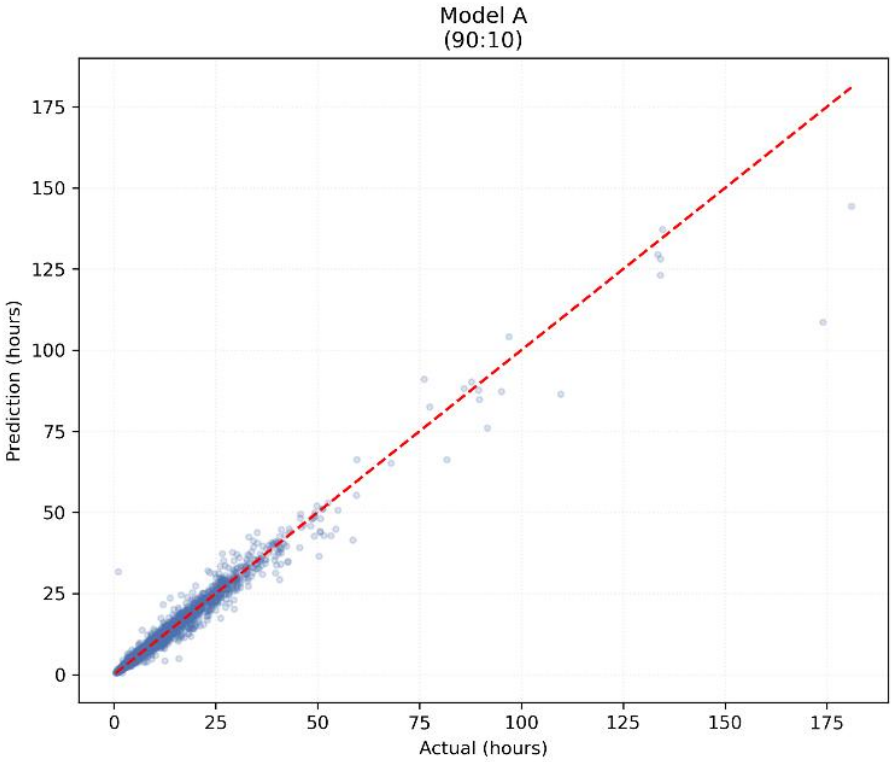
Tabel 4.4 Hasil Evaluasi Semua Model *Split* 90:10

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model A	0.9047	2.2900	0.9570	9.2141
Model B	0.9484	2.3025	0.9565	9.6732
Model C	0.9776	2.4865	0.9493	9.8601
Model D	1.4260	3.1259	0.9198	13.7510

Berdasarkan Tabel 4.4, seluruh model mampu mempertahankan performa yang baik pada skenario *split* 90:10. Model A menghasilkan nilai *error* terendah dibandingkan model lainnya, sedangkan performa menurun secara bertahap pada Model B, Model C, dan Model D seiring berkurangnya informasi fitur yang digunakan. Hasil ini menunjukkan bahwa kelengkapan fitur memberikan pengaruh yang lebih besar terhadap akurasi prediksi dibandingkan perbedaan proporsi data latih dan data uji.

1. Model A

Model A merupakan skenario pemodelan dengan menggunakan seluruh fitur yang tersedia, termasuk fitur karakteristik lintasan (*distance_km, elevation_gain_m, km_effort*) serta fitur lingkungan seperti kondisi cuaca (misalnya suhu dan variabel terkait lainnya). Model ini dirancang sebagai *Baseline* utama dalam penelitian karena merepresentasikan informasi paling lengkap yang dapat digunakan untuk memprediksi *mean finish time*. Dengan kombinasi fitur yang komprehensif, Model A diharapkan mampu menangkap hubungan kompleks antara faktor fisik lintasan dan kondisi lingkungan terhadap performa lomba.



Gambar 4.5 *Parity plot* Model A *Split* 90:10

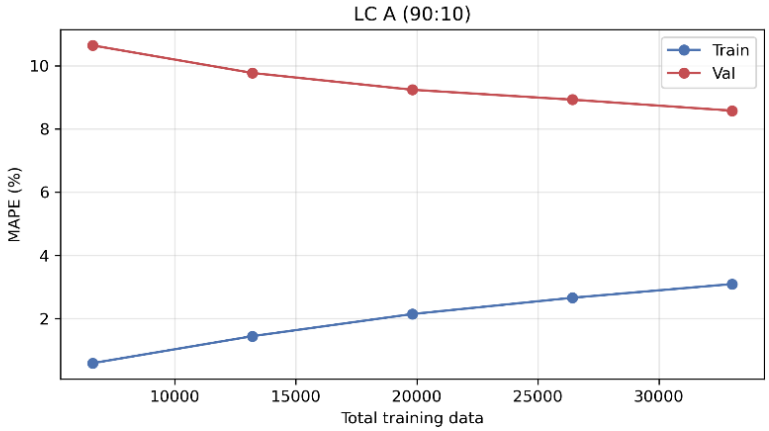
Pada visualisasi *parity plot*, titik-titik prediksi menunjukkan pola yang sangat rapat mengikuti garis diagonal ($y = x$). Hal ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam memetakan hubungan antara fitur *input* dengan target, bahkan pada rentang nilai yang luas. Penyimpangan yang terjadi relatif kecil dan tidak menunjukkan pola sistematis.

Tabel 4.5 Hasil Evaluasi Model A *Split* 90:10

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model A	0.9047	2.2900	0.9570	9.2141

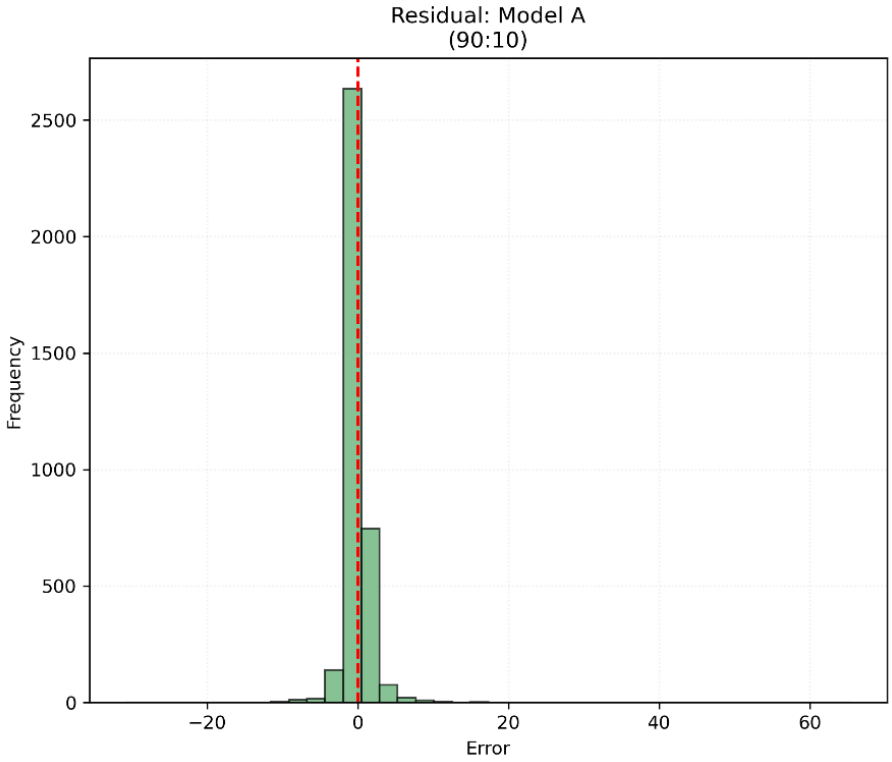
Pada skenario *split* 90:10, Model A menghasilkan MAE sebesar 0,9047 jam, RMSE sebesar 2,2900, R^2 sebesar 0,9570, dan MAPE sebesar 9,2141%. Nilai R^2 yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sekitar 95,7%

variasi *mean finish time*. Tingkat kesalahan relatif yang berada di sekitar 9% menunjukkan bahwa model memiliki kemampuan prediksi yang baik terhadap karakteristik *event Trail Running* yang beragam.



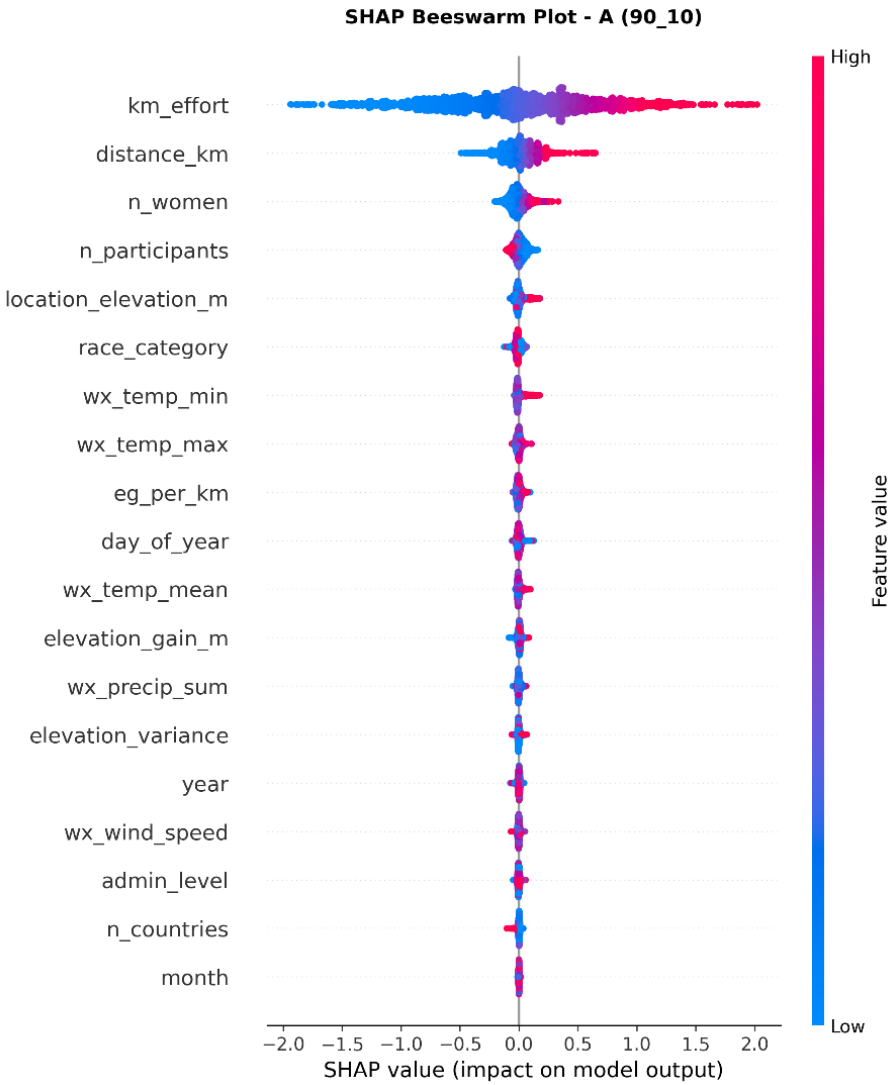
Gambar 4.6 *Learning curve* Model A *Split* 90:10

Kurva pembelajaran menunjukkan bahwa training *error* dan *validation error* mengalami konvergensi pada kisaran nilai yang hampir sama (sekitar 9% MAPE), yang mengindikasikan tidak adanya *overfitting*. Gap yang sangat kecil antara kedua kurva menunjukkan bahwa model memiliki keseimbangan yang baik antara bias dan *variance*.



Gambar 4.7 Residual *Error* Model A *Split* 90:10

Distribusi residual menunjukkan pola yang simetris dan terpusat di sekitar nol, dengan sebagian besar *error* berada dalam rentang kecil (< ±1 jam). Hal ini menandakan bahwa model tidak memiliki bias sistematis dan mampu menghasilkan prediksi yang stabil.

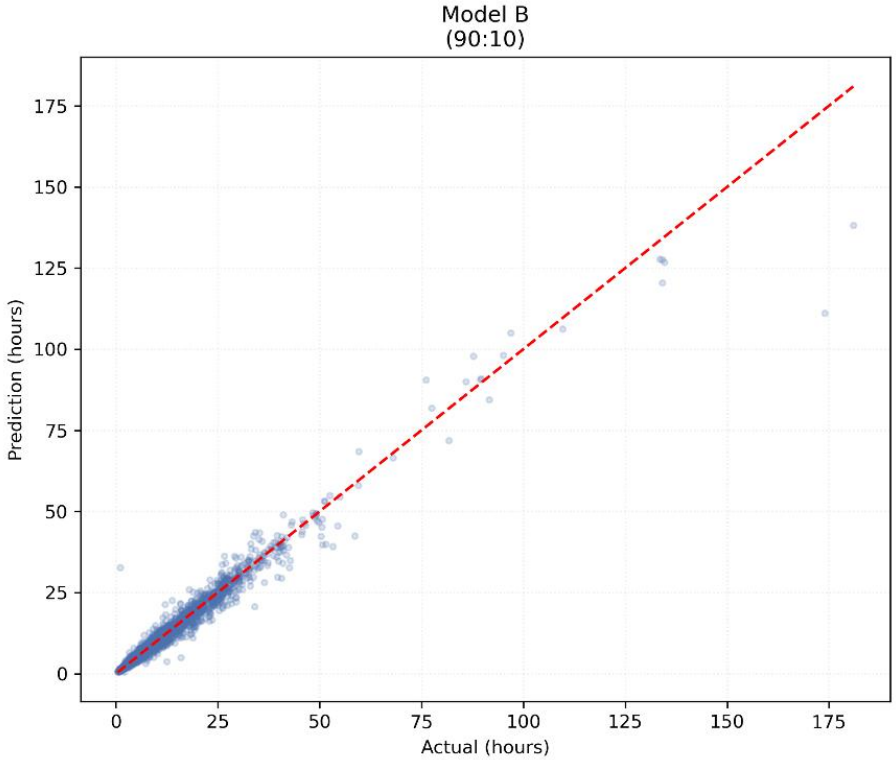


Gambar 4.8 SHAP Analisis Model A *Split* 90:10

Analisis SHAP menunjukkan bahwa fitur *distance_km* dan *elevation_gain_m* merupakan kontributor utama dalam prediksi. Selain itu, fitur cuaca seperti suhu juga memiliki pengaruh signifikan, sehingga nilai SHAP positif pada suhu tinggi berkorelasi dengan peningkatan waktu finis. Hal ini menunjukkan bahwa model mampu menangkap hubungan kontekstual yang kompleks antara kondisi lingkungan dan performa lomba.

2. Model B

Model B merupakan skenario pemodelan yang menggunakan seluruh fitur karakteristik lintasan, namun menghilangkan fitur cuaca. Tujuan dari konfigurasi ini adalah untuk mengevaluasi sejauh mana kontribusi variabel lingkungan terhadap performa model. Dengan membandingkan Model B terhadap Model A, dapat dianalisis dampak langsung dari absennya informasi cuaca terhadap akurasi prediksi.



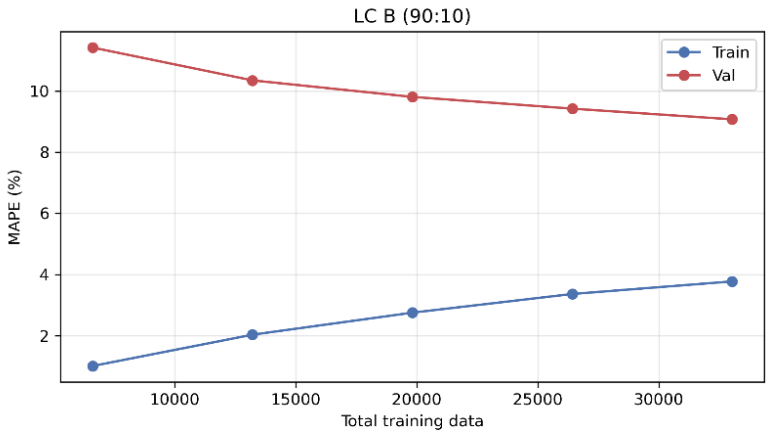
Gambar 4.9 *Parity plot* Model B *Split* 90:10

Dari sisi visualisasi *parity plot*, distribusi titik pada Model B masih mengikuti garis diagonal ($y = x$), namun dengan penyebaran yang lebih luas dibandingkan Model A. Hal ini mencerminkan peningkatan deviasi prediksi, terutama pada nilai ekstrem, yang menunjukkan bahwa model menjadi kurang presisi dalam memodelkan hubungan *non-linear* tanpa informasi cuaca.

Tabel 4.6 Hasil Evaluasi Model B *Split* 90:10

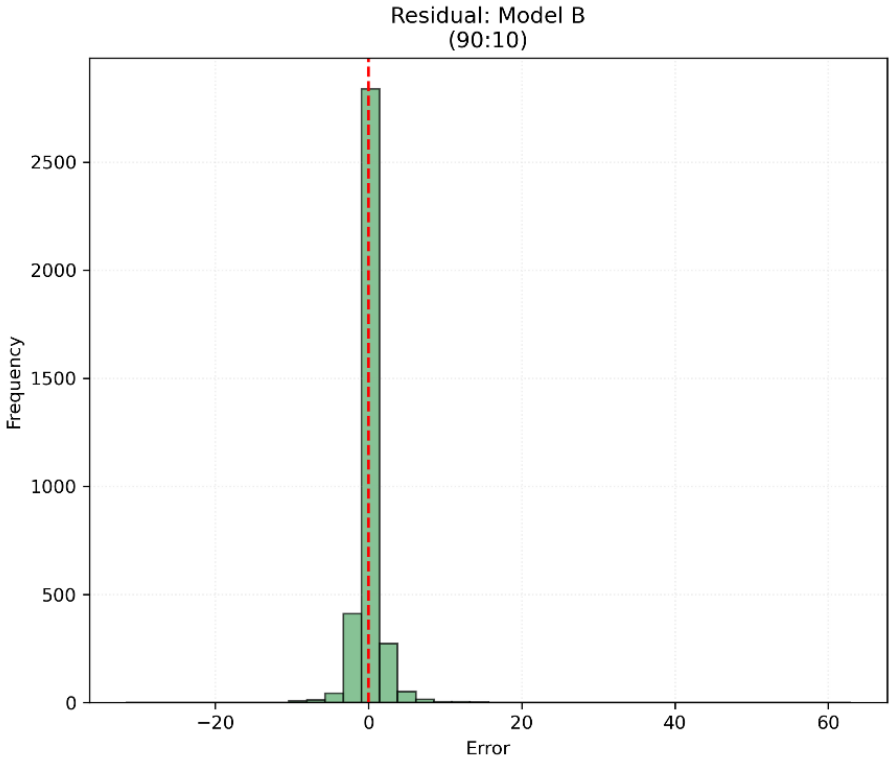
Model	MAE	RMSE	R2	MAPE (%)
Model B	0.9484	2.3025	0.9565	9.6732

Model B menghasilkan MAPE sebesar 9,6732%, sedikit lebih tinggi dibandingkan Model A yang memperoleh MAPE 9,2141%. Selisih sekitar 0,46 poin persentase menunjukkan bahwa informasi cuaca memberikan tambahan informasi yang bermanfaat bagi model. Namun, besarnya perbedaan tersebut relatif kecil sehingga fitur cuaca lebih tepat diposisikan sebagai fitur pendukung yang meningkatkan akurasi secara moderat, bukan sebagai faktor dominan dalam prediksi *mean finish time*.



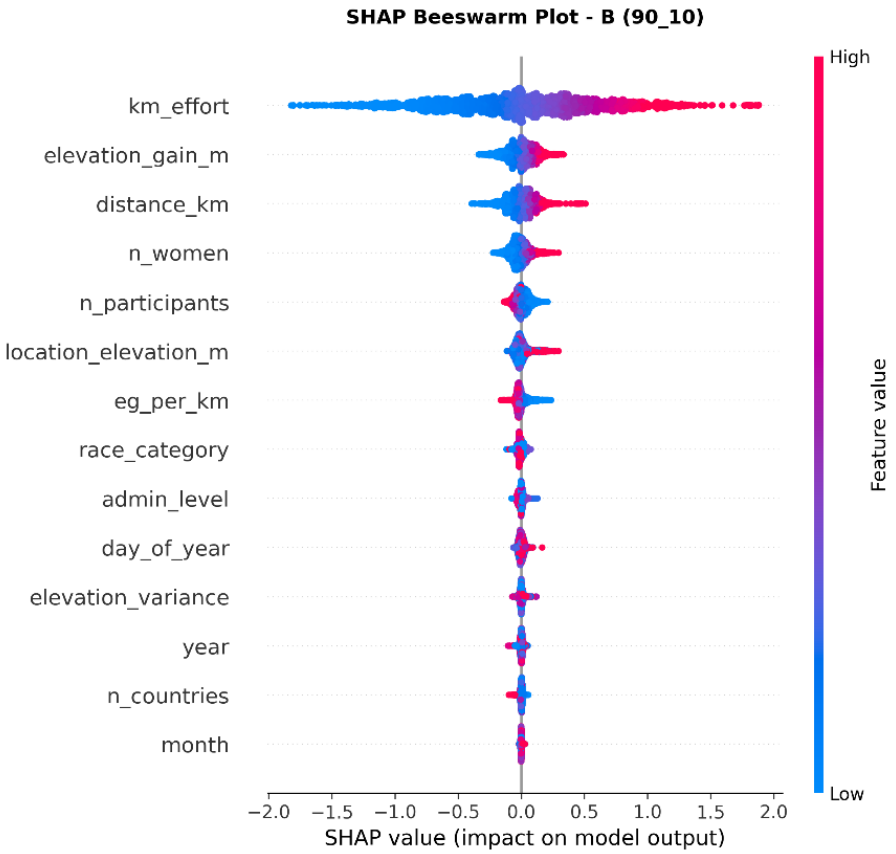
Gambar 4.10 *Learning curve* Model B *Split* 90:10

Kurva pembelajaran (*learning curve*) memperlihatkan bahwa baik training *error* maupun *validation error* tetap mengalami konvergensi, namun berada pada *level error* yang lebih tinggi dibandingkan Model A. Hal ini menunjukkan bahwa meskipun model tidak mengalami *overfitting*, kapasitas generalisasinya menurun akibat keterbatasan informasi fitur.



Gambar 4.11 Residual *Error* Model B *Split* 90:10

Dari sisi distribusi residual, Model B memperlihatkan pola yang masih relatif simetris di sekitar nol, namun dengan variansi yang lebih besar dibandingkan Model A. Hal ini menunjukkan bahwa meskipun model tidak mengalami bias sistematis, tingkat ketidakpastian prediksi meningkat.



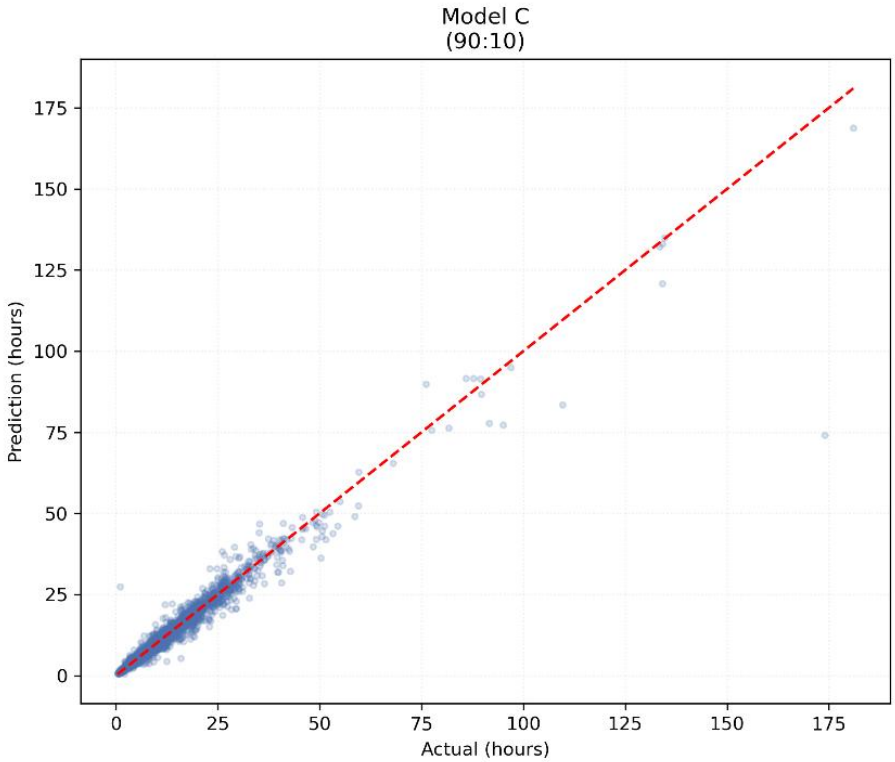
Gambar 4.12 SHAP Analisis Model B *Split* 90:10

Tanpa fitur cuaca, model kehilangan informasi lingkungan → akurasi turun signifikan (~+1.7%). Dengan demikian, jika dibandingkan dengan Model A sebagai *Baseline*, dapat disimpulkan bahwa fitur cuaca memberikan kontribusi signifikan dalam meningkatkan akurasi model.

3. Model C

Model C merupakan skenario pemodelan yang hanya menggunakan sepuluh fitur teratas berdasarkan tingkat kepentingan fitur (*Feature importance*) yang diperoleh dari proses sebelumnya. Pendekatan ini bertujuan untuk menguji apakah model tetap dapat mempertahankan performa yang baik dengan jumlah fitur yang

lebih sedikit, sekaligus meningkatkan efisiensi komputasi. Model ini merepresentasikan pendekatan reduksi dimensi berbasis seleksi fitur.



Gambar 4.13 *Parity plot* Model C *Split* 90:10

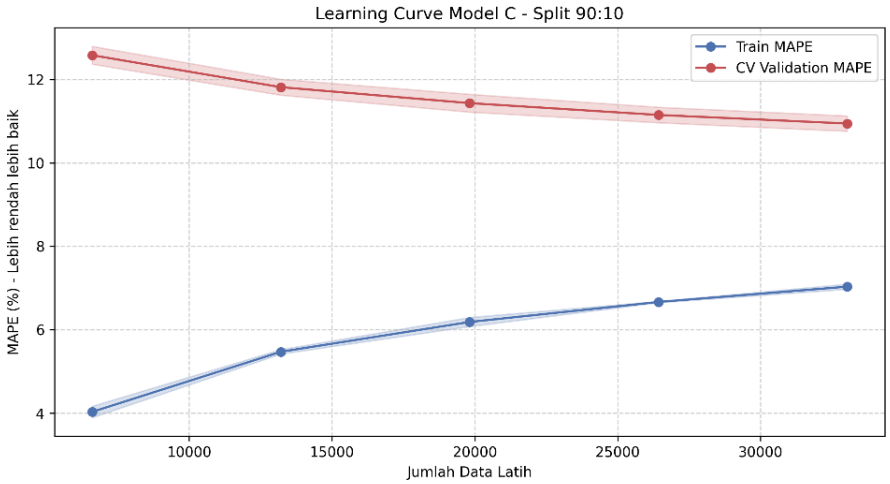
Visualisasi *parity plot* menunjukkan penyebaran titik yang semakin melebar dibandingkan Model B dan Model A, khususnya pada nilai waktu finis yang tinggi. Hal ini mengindikasikan bahwa reduksi fitur berdampak pada kemampuan model dalam menangkap kompleksitas hubungan *non-linear* antar variabel.

Tabel 4.7 Hasil Evaluasi Model C *Split* 90:10

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model C	0.9776	2.4865	0.9493	9.8601

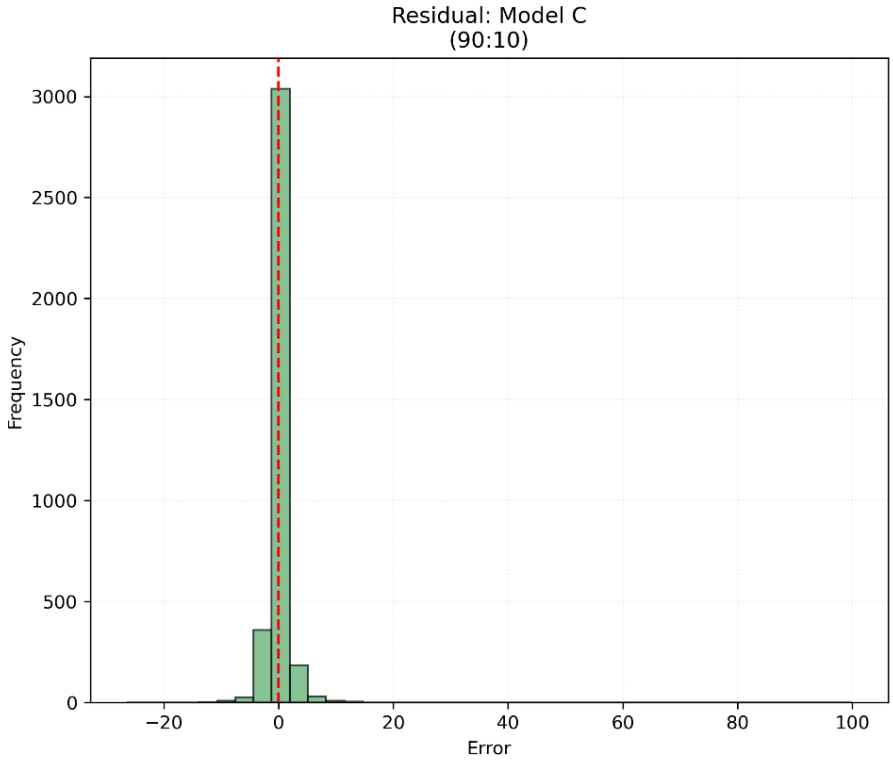
Model C menghasilkan MAPE sebesar 9,8601% dan R^2 sebesar 0,9493. Dibandingkan Model A, terjadi peningkatan *error* yang menunjukkan bahwa proses

reduksi fitur mengurangi sebagian informasi yang diperlukan model untuk menangkap hubungan kompleks antar variabel. Meskipun demikian, performa yang tetap berada di bawah 10% MAPE menunjukkan bahwa sepuluh fitur teratas masih mampu mempertahankan sebagian besar kemampuan prediktif model.



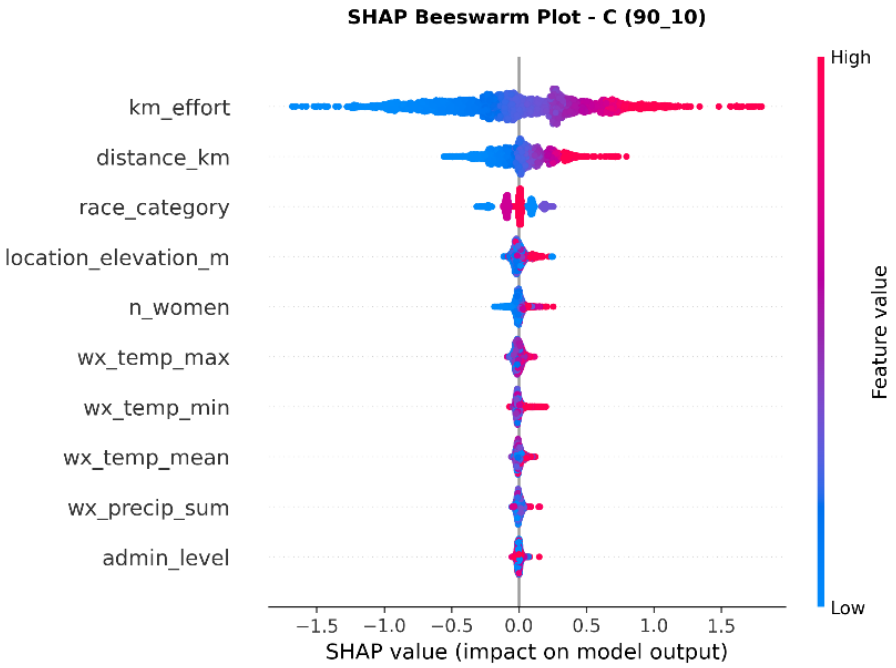
Gambar 4.14 *Learning curve* Model C *Split* 90:10

Kurva pembelajaran menunjukkan pola konvergensi yang stabil antara training dan *validation error*, namun dengan gap yang sedikit lebih besar dibandingkan Model A. Kondisi ini mengindikasikan bahwa meskipun model masih mampu melakukan generalisasi, tingkat akurasinya lebih rendah akibat keterbatasan representasi fitur.



Gambar 4.15 Residual *Error* Model C *Split* 90:10

Distribusi residual tetap terpusat di sekitar nol, tetapi dengan penyebaran yang lebih luas dibandingkan Model A. Hal ini menunjukkan adanya peningkatan *error* yang konsisten di seluruh rentang data.

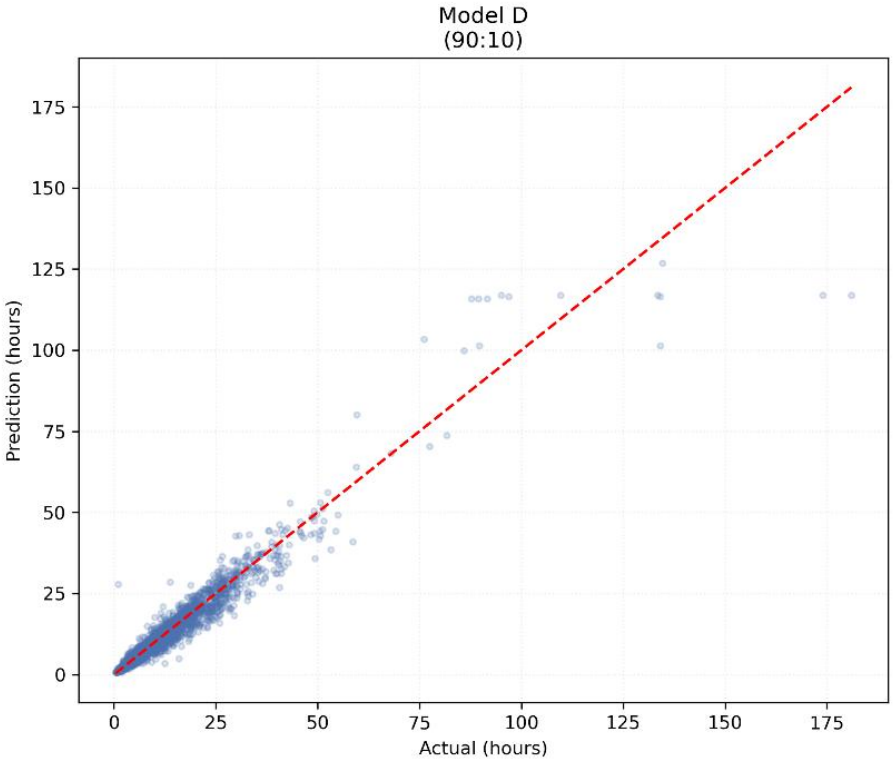


Gambar 4.16 SHAP Analisis Model C *Split* 90:10

Analisis SHAP pada Model C menunjukkan bahwa fitur-fitur utama tetap didominasi oleh variabel terkait lintasan seperti *km_effort* dan *distance_km*. Namun, dibandingkan Model A, kontribusi fitur menjadi lebih terfokus pada beberapa variabel saja, yang mengindikasikan berkurangnya keragaman informasi yang digunakan model dalam melakukan prediksi.

4. Model D

Model D merupakan skenario pemodelan dengan menggunakan fitur minimal, yaitu hanya fitur-fitur dasar yang paling fundamental seperti *distance_km* dan *elevation_gain_m*. Model ini bertujuan untuk mengevaluasi batas bawah performa model ketika informasi yang digunakan sangat terbatas. Dengan demikian, Model D dapat digunakan sebagai pembanding untuk melihat sejauh mana kompleksitas fitur memengaruhi akurasi prediksi.



Gambar 4.17 *Parity plot* Model D *Split* 90:10

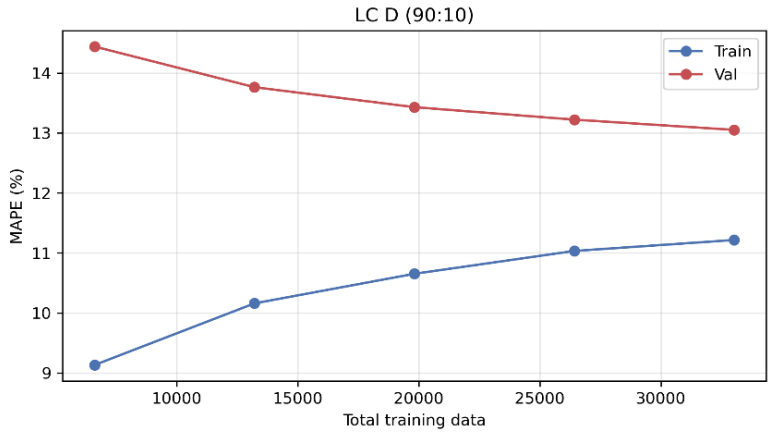
Pada *parity plot*, titik-titik prediksi menunjukkan penyebaran yang jauh lebih luas dari garis diagonal dibandingkan Model A, yang mengindikasikan tingginya tingkat *error* prediksi. Penyimpangan ini terutama terlihat pada nilai ekstrem, yang menunjukkan ketidakmampuan model dalam menangkap variasi data secara menyeluruh.

Tabel 4.8 Hasil Evaluasi Model D *Split* 90:10

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model D	1.4260	3.1259	0.9198	13.7510

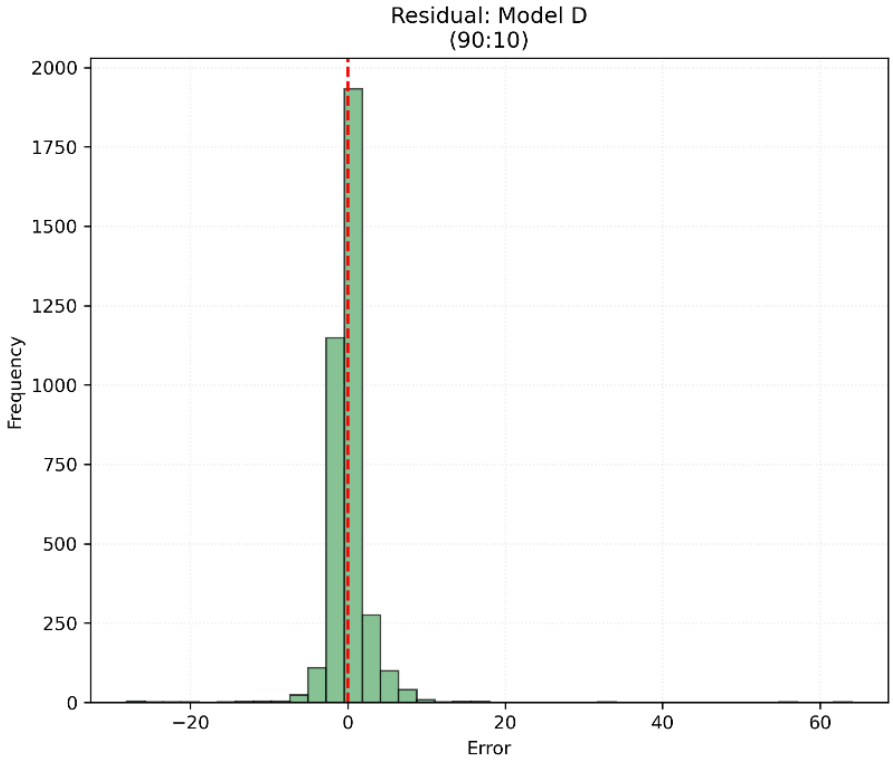
Model D menghasilkan MAPE sebesar 13,7510% dan R^2 sebesar 0,9198. Nilai ini merupakan performa terendah di antara seluruh model yang diuji. Hasil tersebut menunjukkan bahwa penggunaan fitur dasar saja belum cukup untuk merepresentasikan kompleksitas karakteristik *event Trail Running*. Informasi

tambahan di luar jarak dan elevasi masih diperlukan untuk meningkatkan akurasi prediksi.



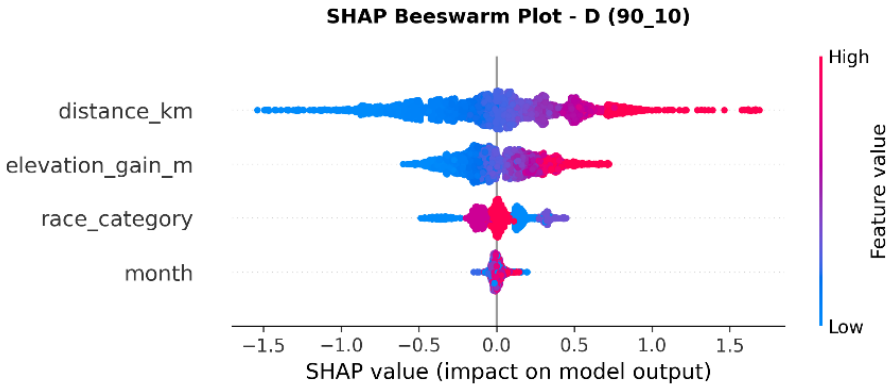
Gambar 4.18 *Learning curve* Model D *Split* 90:10

Kurva pembelajaran menunjukkan konvergensi yang relatif stabil, namun dengan nilai *error* yang jauh lebih tinggi dibandingkan Model A. Hal ini mengindikasikan bahwa keterbatasan fitur menyebabkan model tidak mampu meningkatkan performa meskipun jumlah data latih cukup besar.



Gambar 4.19 Residual *Error* Model D *Split* 90:10

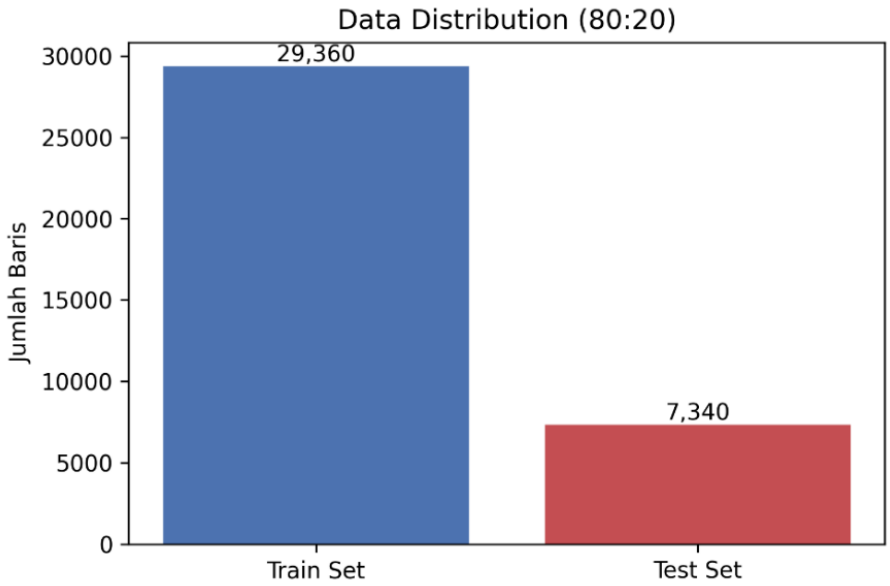
Distribusi residual menunjukkan penyebaran yang jauh lebih luas dibandingkan Model A, yang menandakan peningkatan variansi *error* secara signifikan. Hal ini mencerminkan rendahnya stabilitas model dalam melakukan prediksi.



Gambar 4.20 SHAP *Analisis Model D Split 90:10*

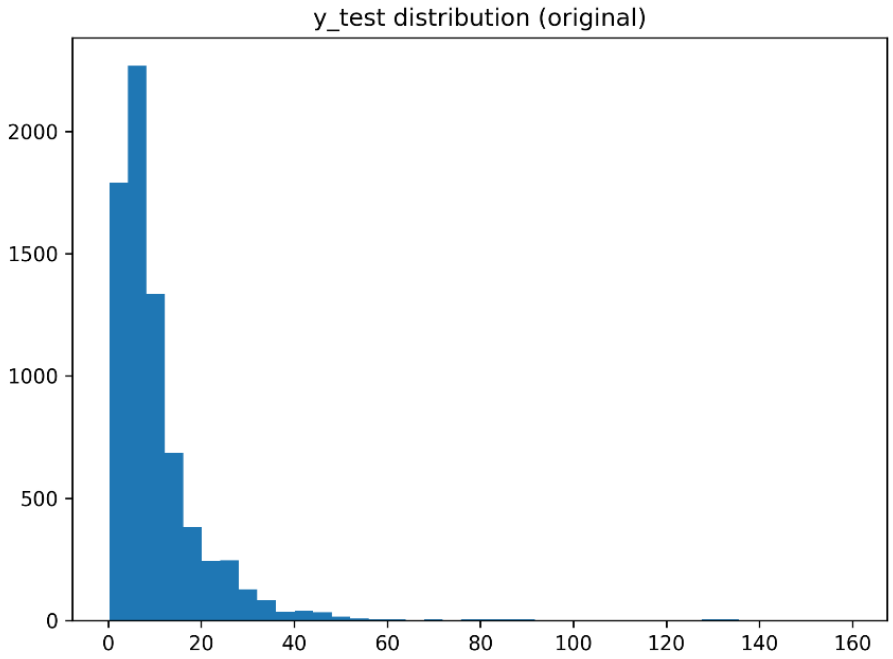
Analisis SHAP memperlihatkan bahwa kontribusi fitur sangat terbatas dan didominasi oleh variabel dasar seperti *distance_km* dan *elevation_gain_m*. Tidak adanya fitur tambahan menyebabkan model kehilangan konteks penting, sehingga akurasi prediksi menurun secara drastis.

4.1.4 Hasil Evaluasi *Split* Data 80 : 20



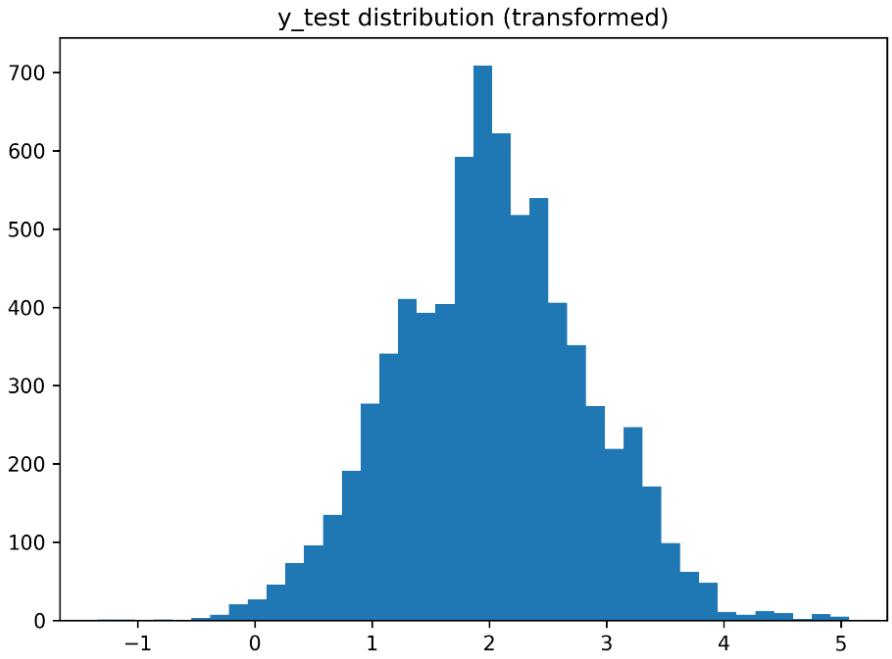
Gambar 4.21 Distribusi *Split* Data 80:20

Pembagian data yang lebih seimbang antara data latih dan uji memberikan evaluasi yang lebih ketat terhadap model, namun sedikit mengurangi jumlah data yang tersedia untuk proses pembelajaran.



Gambar 4.22 Y *Test Distribution* 80:20 Original

Pada skenario *split* 80:20, jumlah observasi dalam data uji meningkat menjadi 7.340 baris. Meskipun jumlah sampel lebih banyak, Gambar 4.22 menunjukkan bahwa pola distribusi *mean finish time* tetap konsisten dengan skenario 90:10, sehingga frekuensi tertinggi tetap berada pada durasi di bawah 25 jam. Hal ini membuktikan bahwa pengambilan sampel acak yang dilakukan tetap representatif terhadap karakteristik populasi asli.



Gambar 4.23 Y *Test Distribution* 80:20 *Transformed*

Sama halnya dengan skenario sebelumnya, transformasi logaritmik pada Gambar 4.23 berhasil menarik nilai-nilai ekstrem ke arah pusat distribusi. Dengan sebaran yang lebih merata ini, model dapat mempelajari variansi data secara lebih adil di seluruh rentang waktu, baik untuk lomba jarak pendek maupun jarak jauh.

Tabel 4.9 Hasil Evaluasi Semua Model *Split* Data 80:20

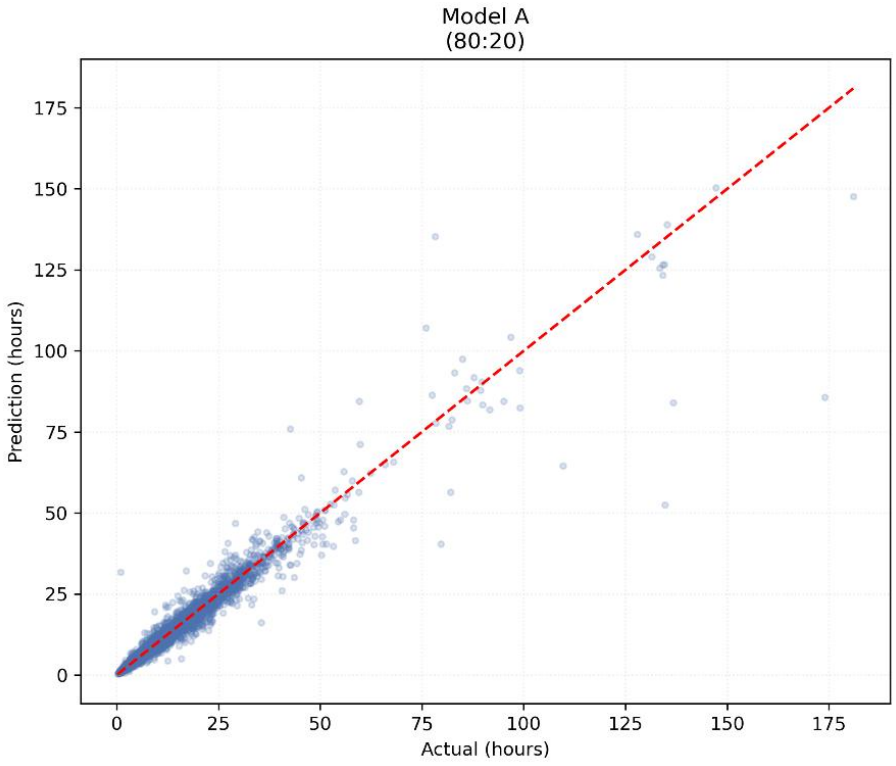
Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model A	0.9267	2.5732	0.9452	8.9117
Model B	0.9427	2.3236	0.9553	9.3195
Model C	0.9859	2.5779	0.9450	9.5688
Model D	1.3564	2.8404	0.9333	13.4438

Berdasarkan Tabel 4.9, pola performa seluruh model tetap konsisten dengan skenario sebelumnya. Model A tetap menjadi model dengan *error* terendah, diikuti oleh Model B, Model C, dan Model D. Konsistensi urutan tersebut menunjukkan

bahwa pengaruh komposisi fitur terhadap performa model lebih dominan dibandingkan perubahan proporsi pembagian data.

1. Model A

Model A merupakan skenario pemodelan dengan menggunakan seluruh fitur yang tersedia, termasuk fitur karakteristik lintasan (*distance_km, elevation_gain_m, km_effort*) serta fitur lingkungan seperti kondisi cuaca (misalnya suhu dan variabel terkait lainnya). Model ini dirancang sebagai *Baseline* utama dalam penelitian karena merepresentasikan informasi paling lengkap yang dapat digunakan untuk memprediksi *mean finish time*. Dengan kombinasi fitur yang komprehensif, Model A diharapkan mampu menangkap hubungan kompleks antara faktor fisik lintasan dan kondisi lingkungan terhadap performa lomba.



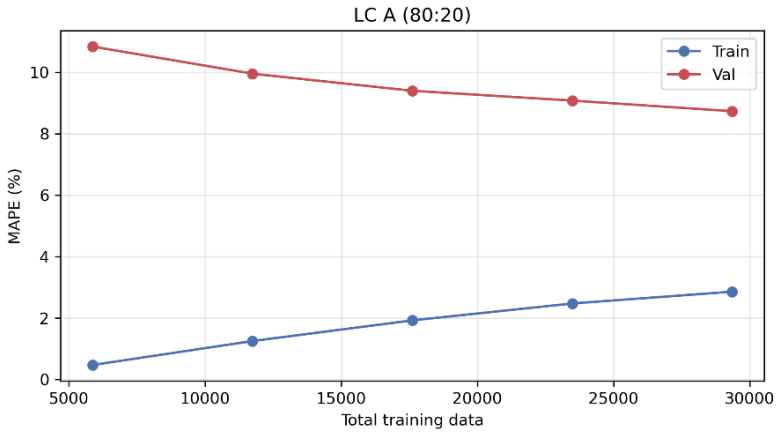
Gambar 4.24 *Parity plot* Model A *Split* 80:20

Parity plot pada skenario ini masih menunjukkan pola yang mengikuti garis diagonal, namun dengan penyebaran titik yang sedikit lebih lebar dibandingkan *split* 90:10. Hal ini mengindikasikan adanya peningkatan deviasi prediksi, terutama pada nilai ekstrem, meskipun secara keseluruhan model masih mempertahankan akurasi yang tinggi.

Tabel 4.10 Hasil Evaluasi Model A *Split* Data 80:20

Model	MAE	RMSE	R2	MAPE (%)
Model A	0.9267	2.5732	0.9452	8.9117

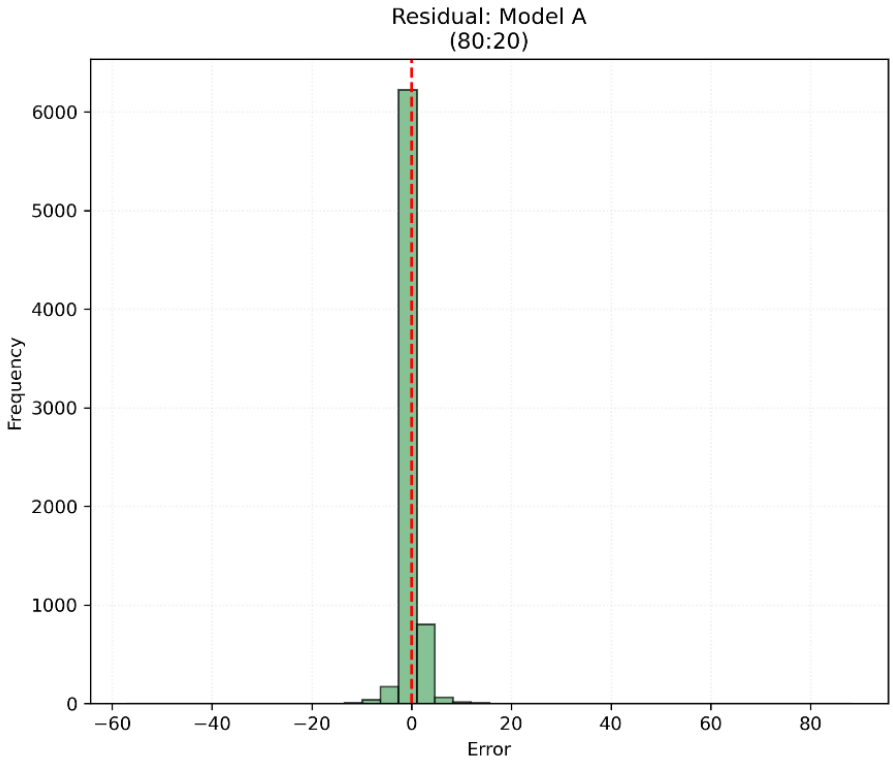
Pada skenario *split* 80:20, Model A menghasilkan MAE sebesar 0,9267 jam, RMSE sebesar 2,5732, R^2 sebesar 0,9452, dan MAPE sebesar 8,9117%. Nilai MAPE ini sedikit lebih rendah dibandingkan skenario 90:10. Perbedaan tersebut menunjukkan bahwa perubahan proporsi data tidak menghasilkan pola performa yang konsisten, sehingga tidak dapat disimpulkan bahwa semakin besar data latih selalu menghasilkan akurasi yang lebih baik.



Gambar 4.25 *Learning curve* Model A *Split* 80:20

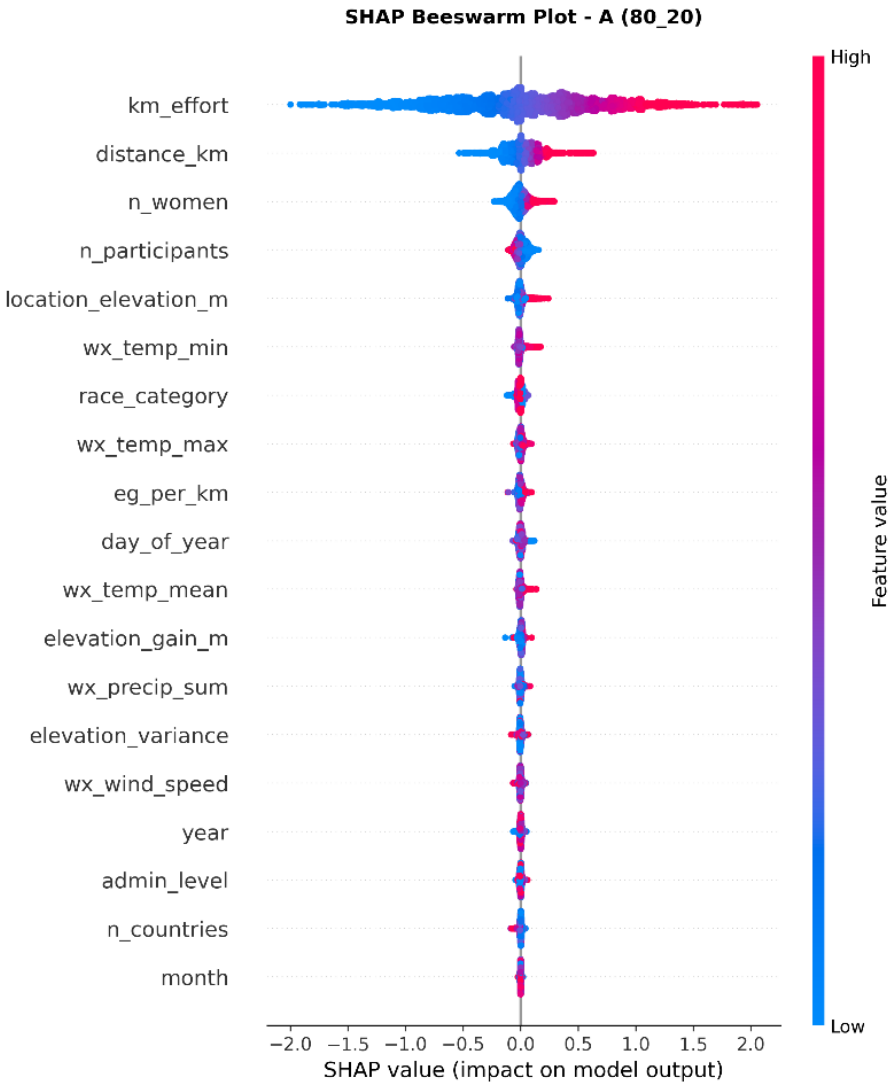
Kurva pembelajaran tetap menunjukkan konvergensi antara training dan *validation error*, namun dengan gap yang sedikit lebih besar dibandingkan *split*

sebelumnya. Hal ini menunjukkan adanya peningkatan *variance* akibat berkurangnya jumlah data latih, meskipun masih dalam batas yang dapat diterima.



Gambar 4.26 Residual *Error* Model A *Split* 80:20

Distribusi residual tetap terpusat di sekitar nol, namun dengan penyebaran yang lebih luas dibandingkan *split* 90:10. Hal ini mencerminkan peningkatan variansi *error*, terutama pada data dengan waktu finis yang tinggi.



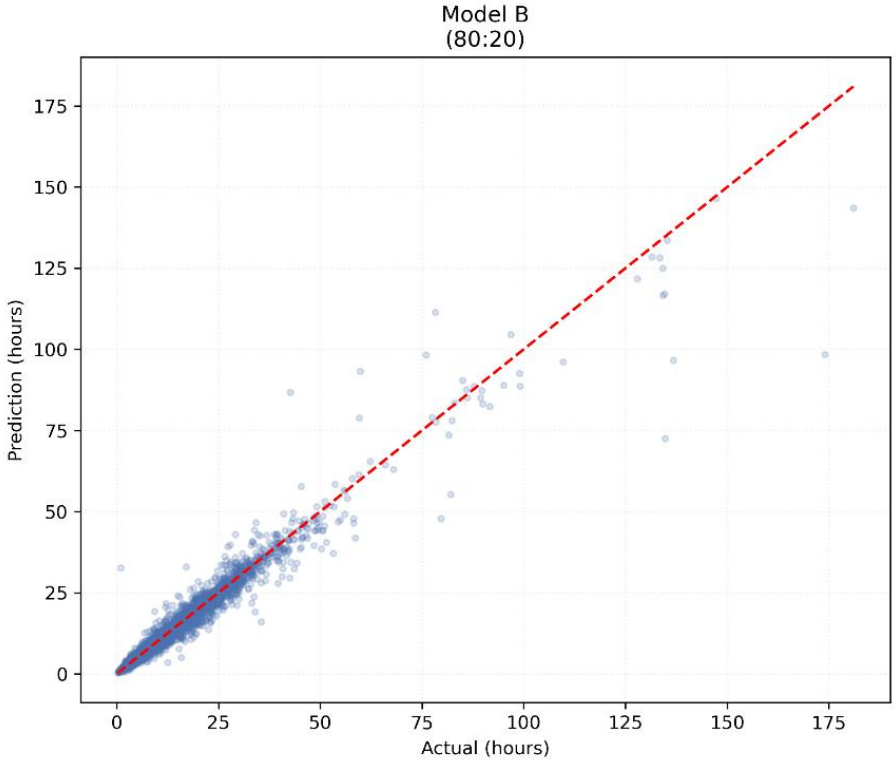
Gambar 4.27 SHAP Analisis Model A *Split* 80:20

Analisis SHAP menunjukkan bahwa urutan kontribusi fitur tetap konsisten dengan *split* sebelumnya, sehingga *distance_km* dan *elevation_gain_m* tetap menjadi faktor dominan. Konsistensi ini menunjukkan bahwa hubungan antar fitur bersifat stabil meskipun proporsi data berubah.

2. Model B

Model B merupakan skenario pemodelan yang menggunakan seluruh fitur karakteristik lintasan, namun menghilangkan fitur cuaca. Tujuan dari konfigurasi

ini adalah untuk mengevaluasi sejauh mana kontribusi variabel lingkungan terhadap performa model. Dengan membandingkan Model B terhadap Model A, dapat menganalisis dampak langsung dari absennya informasi cuaca terhadap akurasi prediksi.



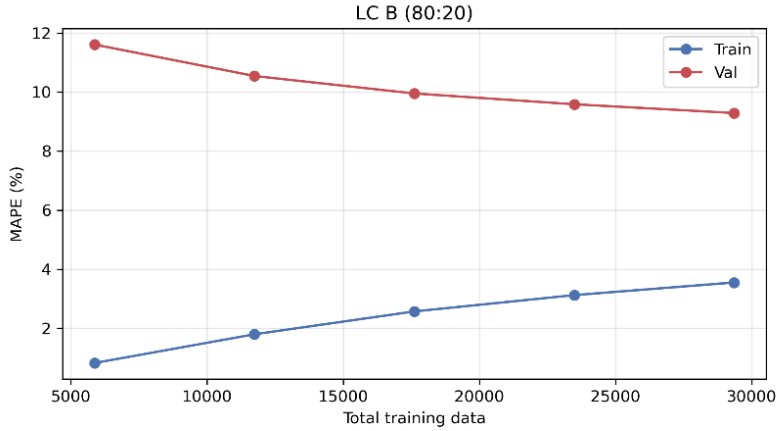
Gambar 4.28 *Parity plot* Model B *Split* 80:20

Pada visualisasi *parity plot*, titik-titik prediksi pada Model B masih mengikuti pola diagonal, namun menunjukkan penyebaran yang lebih luas dibandingkan Model A, terutama pada rentang waktu finis yang tinggi (>50 jam). Hal ini mengindikasikan bahwa model kehilangan sensitivitas terhadap kondisi ekstrem ketika variabel lingkungan tidak disertakan.

Tabel 4.11 Hasil Evaluasi Model B *Split* Data 80:20

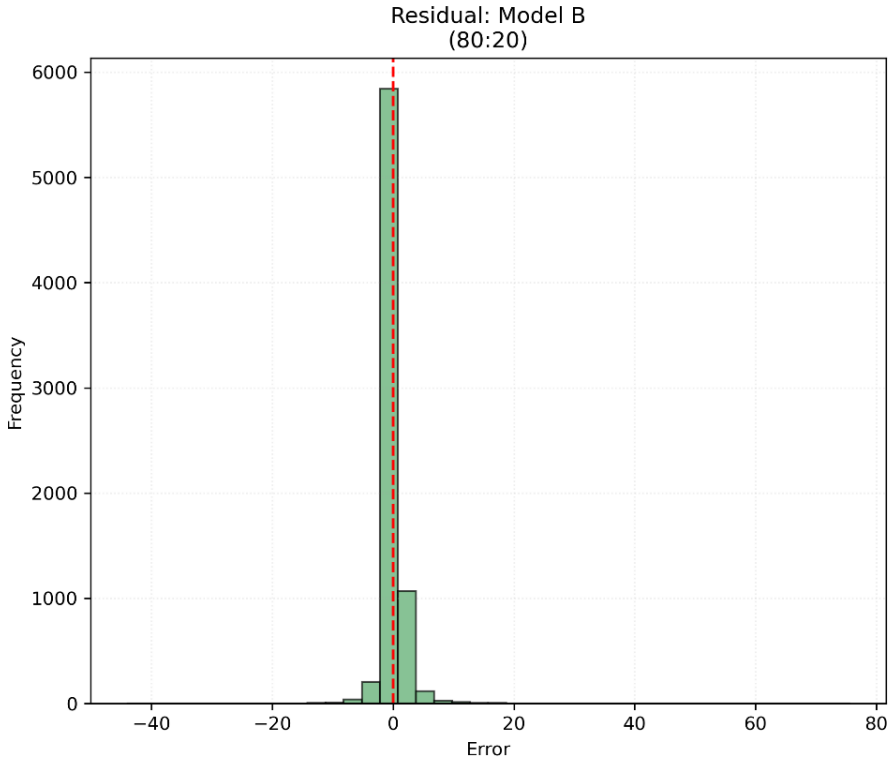
Model	MAE	RMSE	R2	MAPE (%)
Model B	0.9427	2.3236	0.9553	9.3195

Model B menghasilkan MAPE sebesar 9,3195%, hanya berbeda sekitar 0,41 poin persentase dibandingkan Model A. Hasil ini menunjukkan bahwa model tetap mampu mempertahankan sebagian besar kemampuan prediksi meskipun seluruh fitur cuaca dihilangkan. Dengan demikian, Model B dapat dipertimbangkan sebagai alternatif ketika data cuaca tidak tersedia.



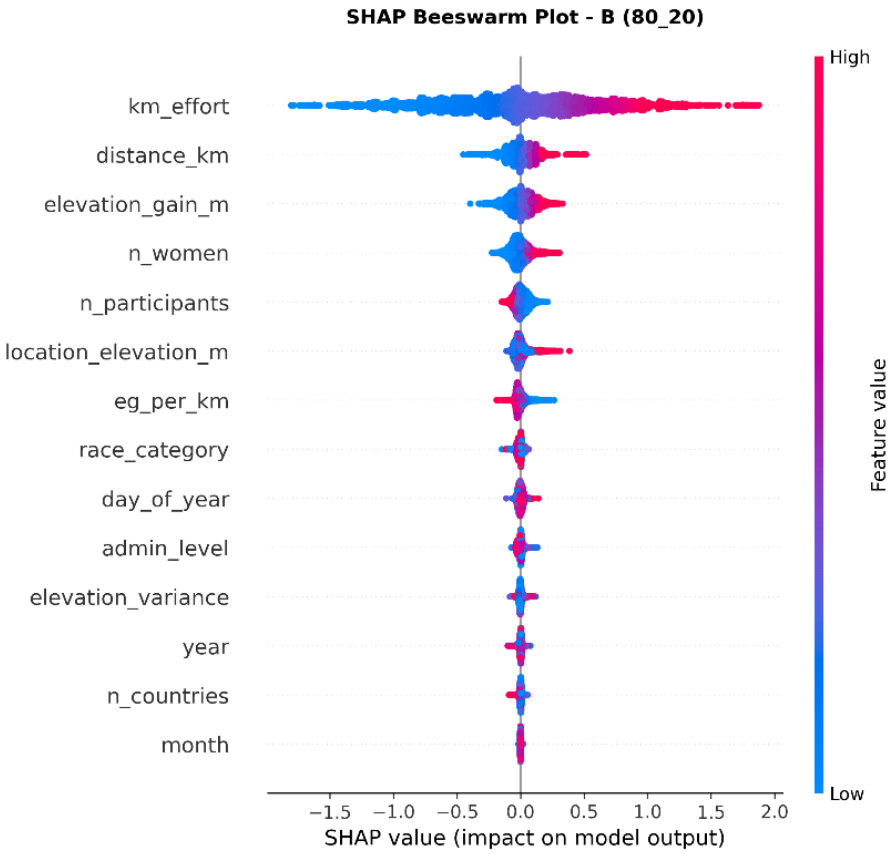
Gambar 4.29 *Learning curve* Model B *Split* 80:20

Kurva pembelajaran menunjukkan konvergensi yang tetap terjaga, namun dengan gap yang sedikit lebih besar dibandingkan Model A. Hal ini menunjukkan adanya penurunan stabilitas generalisasi akibat berkurangnya informasi fitur.



Gambar 4.30 Residual *Error* Model B *Split* 80:20

Distribusi residual menunjukkan pola yang masih acak, tetapi dengan variansi yang lebih tinggi dibandingkan Model A. Hal ini menegaskan bahwa model menjadi kurang stabil dalam menghasilkan prediksi.



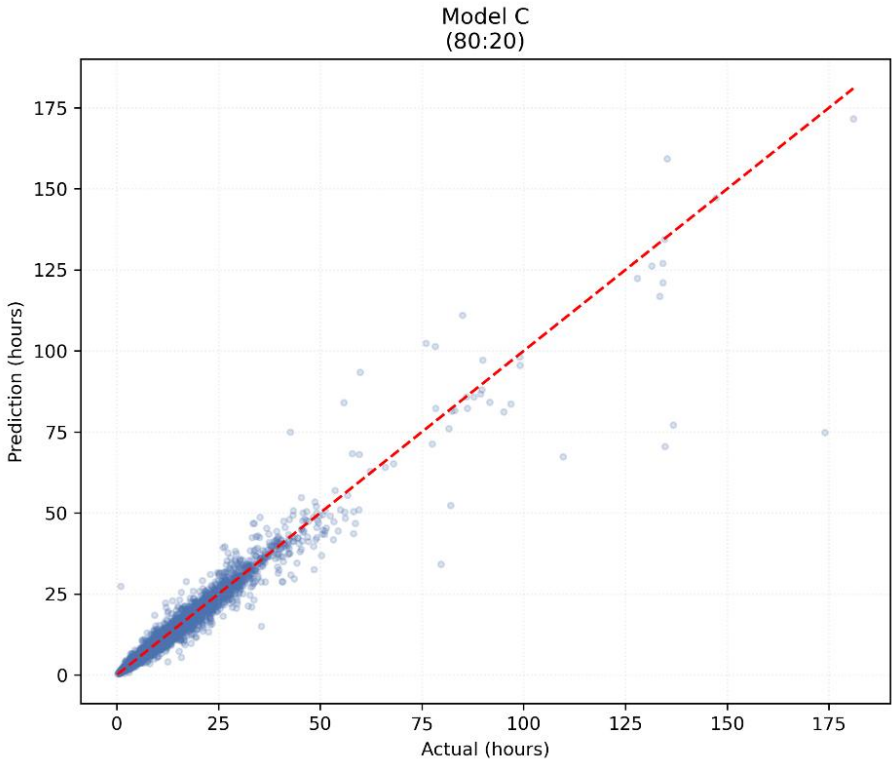
Gambar 4.31 SHAP Analisis Model B *Split* 80:20

Analisis SHAP menunjukkan pola kontribusi fitur yang serupa dengan *split* sebelumnya, sehingga fitur lintasan tetap dominan, namun tidak adanya variabel cuaca menyebabkan berkurangnya kedalaman informasi yang digunakan dalam proses prediksi.

3. Model C

Model C merupakan skenario pemodelan yang hanya menggunakan sepuluh fitur teratas berdasarkan tingkat kepentingan fitur (*Feature importance*) yang diperoleh dari proses sebelumnya. Pendekatan ini bertujuan untuk menguji apakah model tetap dapat mempertahankan performa yang baik dengan jumlah fitur yang

lebih sedikit, sekaligus meningkatkan efisiensi komputasi. Model ini merepresentasikan pendekatan reduksi dimensi berbasis seleksi fitur.



Gambar 4.32 *Parity plot* Model C *Split* 80:20

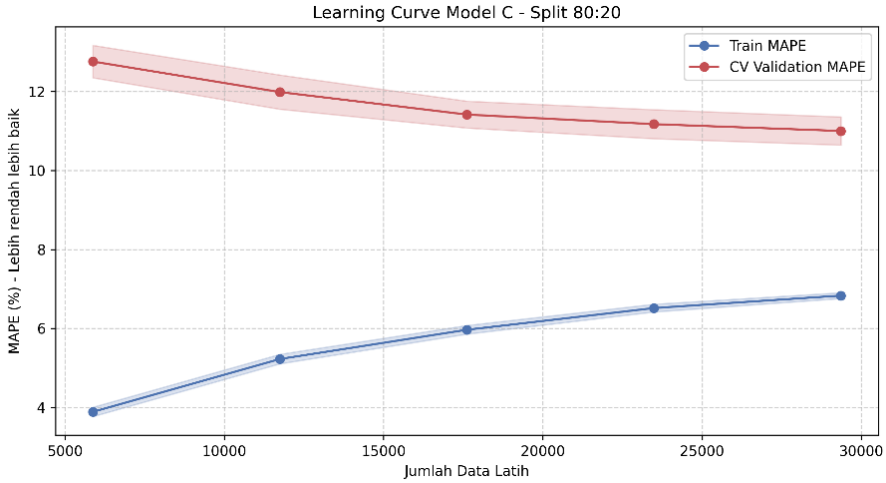
Parity plot memperlihatkan bahwa penyebaran titik semakin melebar dibandingkan Model B dan Model A, khususnya pada nilai prediksi yang tinggi. Hal ini mengindikasikan bahwa model mengalami kesulitan dalam memodelkan hubungan *non-linear* secara optimal ketika jumlah fitur dibatasi.

Tabel 4.12 Hasil Evaluasi Model C *Split* Data 80:20

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model C	0.9859	2.5779	0.9450	9.5688

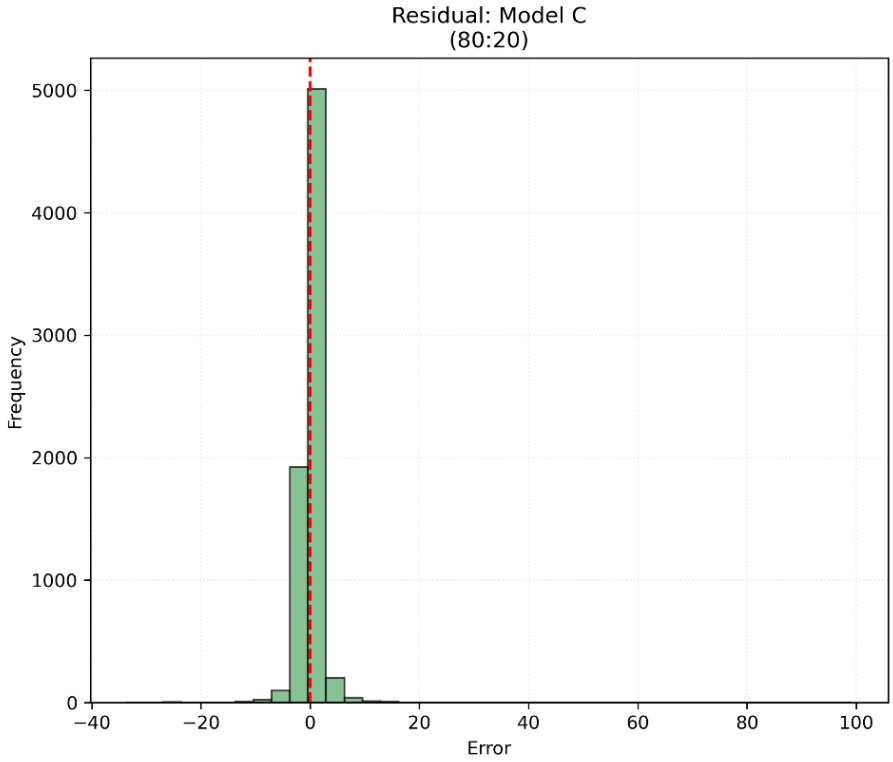
Model C menghasilkan MAPE sebesar 9,5688% dan R^2 sebesar sekitar 0,945. Walaupun performanya sedikit berada di bawah Model A dan Model B, hasil

tersebut menunjukkan bahwa pendekatan seleksi fitur masih mampu mempertahankan akurasi yang baik dengan jumlah fitur yang lebih sedikit.



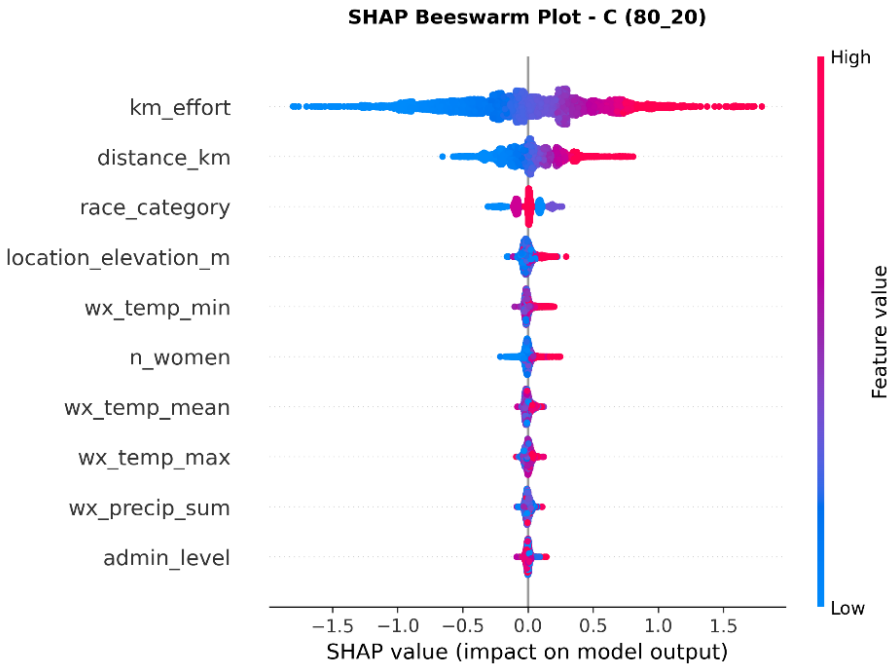
Gambar 4.33 *Learning curve* Model C *Split* 80:20

Kurva pembelajaran menunjukkan konvergensi yang stabil, namun dengan nilai training dan *validation error* yang lebih tinggi dibandingkan Model A. Selain itu, gap yang sedikit lebih besar menunjukkan adanya penurunan efisiensi generalisasi.



Gambar 4.34 Residual *Error* Model C *Split* 80:20

Distribusi residual tetap terpusat di sekitar nol, namun dengan penyebaran yang lebih luas dibandingkan Model A. Hal ini menunjukkan bahwa *error* prediksi meningkat secara sistematis.

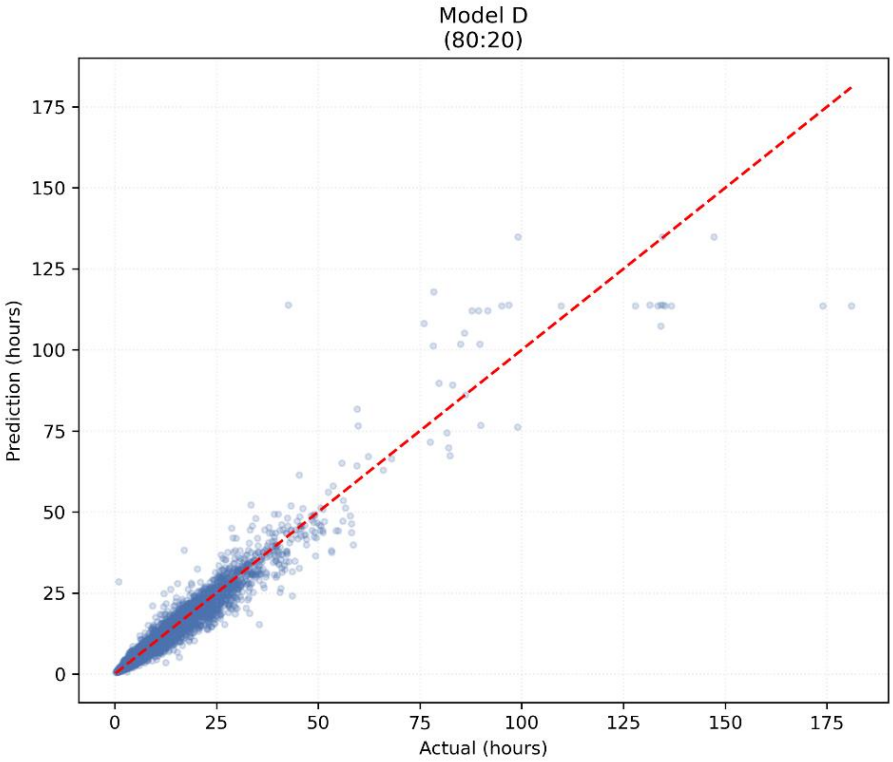


Gambar 4.35 SHAP Analisis Model C *Split* 80:20

Analisis SHAP menunjukkan bahwa fitur utama seperti *km_effort*, *distance_km*, dan *elevation_gain_m* tetap menjadi kontributor utama. Namun, dibandingkan Model A, distribusi kontribusi fitur menjadi lebih terkonsentrasi pada sedikit variabel, yang mengindikasikan berkurangnya kompleksitas representasi model.

4. Model D

Model D merupakan skenario pemodelan dengan menggunakan fitur minimal, yaitu hanya fitur-fitur dasar yang paling fundamental seperti *distance_km* dan *elevation_gain_m*. Model ini bertujuan untuk mengevaluasi batas bawah performa model ketika informasi yang digunakan sangat terbatas. Dengan demikian, Model D dapat digunakan sebagai pembanding untuk melihat sejauh mana kompleksitas fitur memengaruhi akurasi prediksi.



Gambar 4.36 *Parity plot* Model D *Split* 80:20

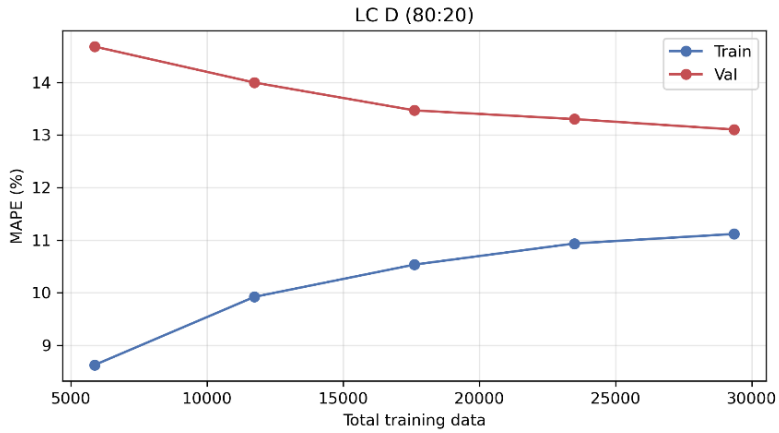
Parity plot menunjukkan penyebaran titik yang jauh lebih luas dari garis diagonal, yang menandakan tingginya deviasi antara nilai prediksi dan aktual. Ketidaktepatan ini semakin terlihat pada nilai ekstrem.

Tabel 4.13 Hasil Evaluasi Model D *Split* Data 80:20

Model	MAE	RMSE	R2	MAPE (%)
Model D	1.3564	2.8404	0.9333	13.4438

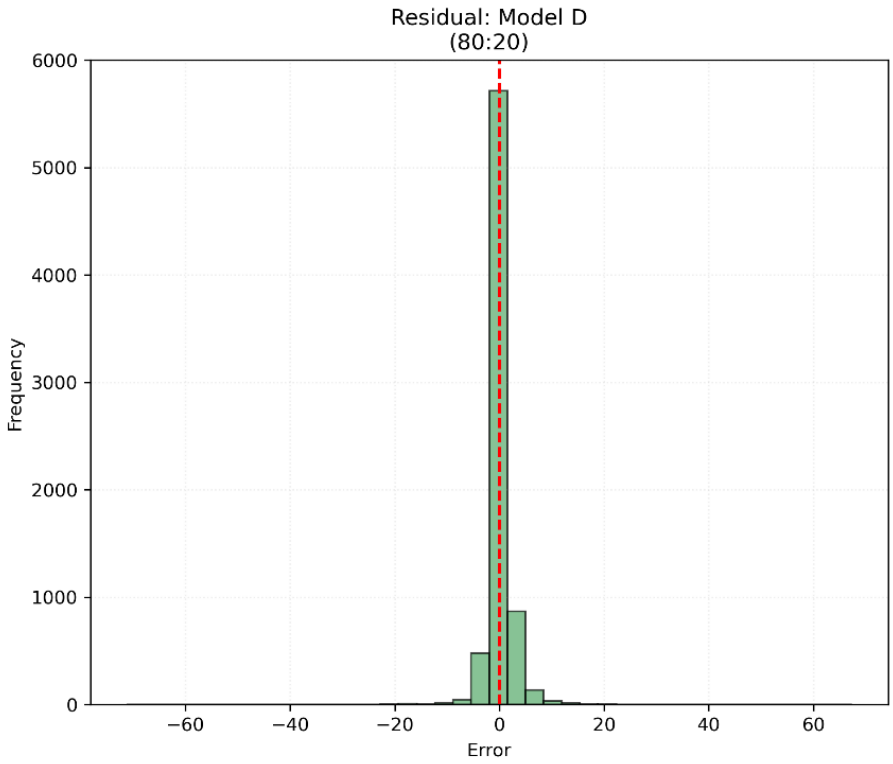
Model D memperoleh nilai MAE sebesar 1,3564 jam, RMSE sebesar 2,8404, R^2 sebesar 0,9333, dan MAPE sebesar 13,4438%. Nilai tersebut merupakan performa terendah dibandingkan model lainnya pada skenario *split* 80:20. Dibandingkan dengan Model A yang menggunakan fitur lengkap, terjadi peningkatan MAPE sebesar sekitar 4,53 poin persentase serta penurunan nilai R^2 sebesar 0,012. Hasil ini menunjukkan bahwa penggunaan fitur dasar saja belum

mampu merepresentasikan kompleksitas faktor yang memengaruhi *mean finish time* secara optimal.



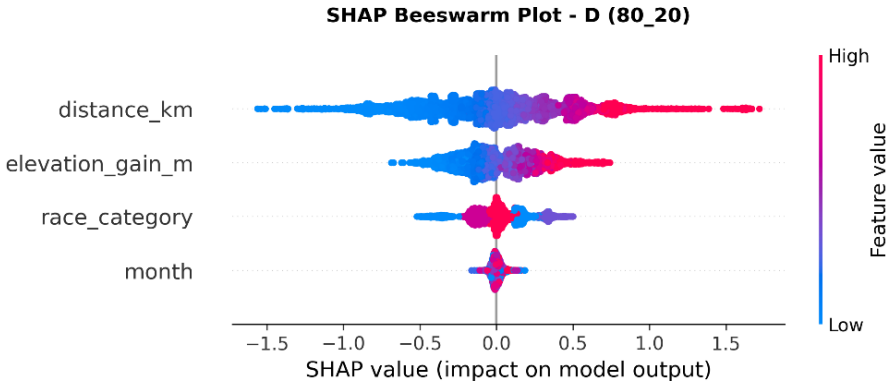
Gambar 4.37 *Learning curve* Model D *Split* 80:20

Kurva pembelajaran menunjukkan konvergensi yang relatif stabil, namun dengan nilai *error* yang jauh lebih tinggi dibandingkan Model A. Hal ini menunjukkan bahwa keterbatasan fitur membatasi kemampuan model dalam meningkatkan performa, meskipun data latih tersedia dalam jumlah yang cukup.



Gambar 4.38 Residual *Error* Model D *Split* 80:20

Distribusi residual menunjukkan variansi yang besar, yang mengindikasikan rendahnya stabilitas model dalam menghasilkan prediksi yang konsisten.

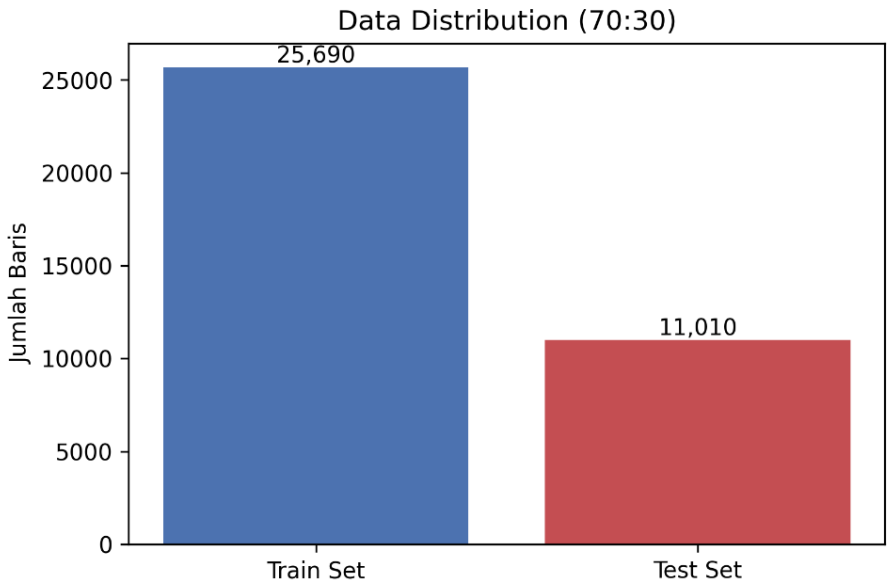


Gambar 4.39 SHAP Analisis Model D *Split* 80:20

Analisis SHAP menunjukkan bahwa kontribusi fitur sangat terbatas dan didominasi oleh variabel dasar seperti *distance_km*. Hal ini menegaskan bahwa

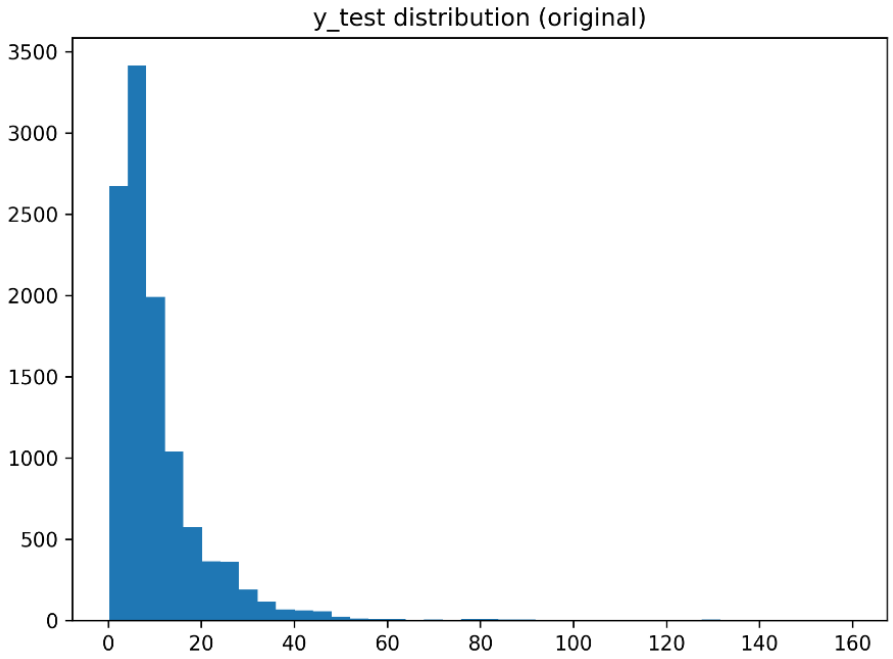
model tidak memiliki cukup informasi untuk menangkap kompleksitas hubungan dalam data.

4.1.5 Hasil Evaluasi *Split* Data 70 : 30



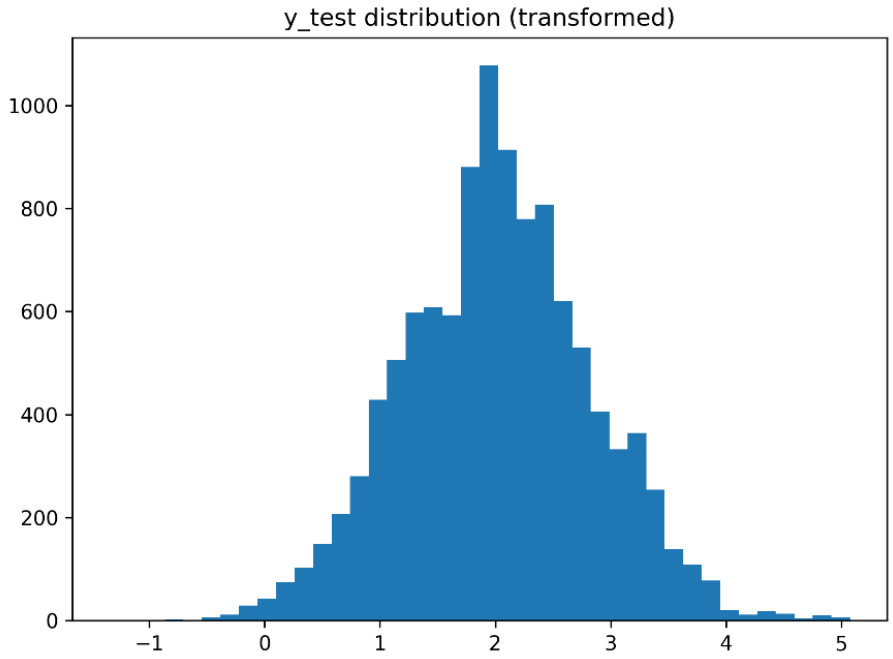
Gambar 4.40 Distribusi *Split* Data 70:30

Proporsi data uji yang lebih besar meningkatkan reliabilitas evaluasi, tetapi berpotensi menurunkan performa model karena keterbatasan data latih.



Gambar 4.41 Y *Test Distribution* 70:30 Original

Gambar 4.41 menyajikan data uji dengan porsi terbesar (30% atau sekitar 11.010 baris). Dengan jumlah data uji yang semakin besar, variansi data terlihat semakin jelas pada area ekor grafik. Grafik ini memberikan tantangan evaluasi yang lebih ketat bagi model karena harus memprediksi sampel yang lebih beragam dengan jumlah data latih yang lebih sedikit (hanya 70%).



Gambar 4.42 Y *Test Distribution* 70:30 *Transformed*

Visualisasi terakhir pada Gambar 4.42 mengonfirmasi bahwa metode transformasi log tetap efektif bahkan saat ukuran data uji diperbesar. Konsistensi bentuk lonceng pada grafik ini menunjukkan bahwa draf *preprocessing* yang dirancang dalam penelitian ini memiliki stabilitas yang baik untuk diterapkan pada berbagai proporsi pembagian data.

Tabel 4.14 Hasil Evaluasi Semua Model *Split* Data 70:30

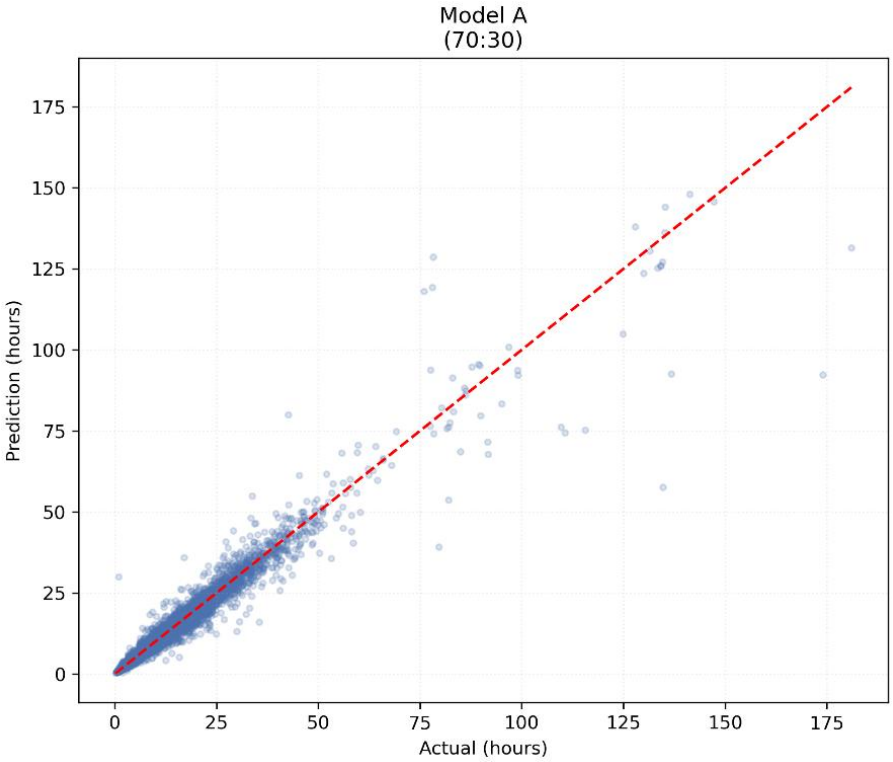
Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model A	0.9351	2.3875	0.9511	8.8529
Model B	0.9517	2.4027	0.9505	9.2586
Model C	0.9986	2.6052	0.9418	9.5464
Model D	1.3500	2.8098	0.9323	13.2360

Berdasarkan Tabel 4.14, urutan performa model pada skenario *split* 70:30 tetap konsisten dengan skenario sebelumnya. Model A menghasilkan performa

terbaik dengan MAPE sebesar 8.8529%, diikuti oleh Model B sebesar 9.2586%, Model C sebesar 9.5464%, dan Model D sebesar 13.2360%. Peningkatan *error* yang terjadi secara bertahap dari Model A hingga Model D menunjukkan bahwa semakin sedikit informasi yang diberikan kepada model, semakin rendah pula kemampuan model dalam menjelaskan variasi *mean finish time*. Konsistensi pola ini mengindikasikan bahwa pengaruh komposisi fitur terhadap performa model lebih dominan dibandingkan perubahan proporsi data latih dan data uji.

1. Model A

Model A merupakan skenario pemodelan dengan menggunakan seluruh fitur yang tersedia, termasuk fitur karakteristik lintasan (*distance_km, elevation_gain_m, km_effort*) serta fitur lingkungan seperti kondisi cuaca (misalnya suhu dan variabel terkait lainnya). Model ini dirancang sebagai *Baseline* utama dalam penelitian karena merepresentasikan informasi paling lengkap yang dapat digunakan untuk memprediksi *mean finish time*. Dengan kombinasi fitur yang komprehensif, Model A diharapkan mampu menangkap hubungan kompleks antara faktor fisik lintasan dan kondisi lingkungan terhadap performa lomba.



Gambar 4.43 *Parity plot* Model A *Split* 70:30

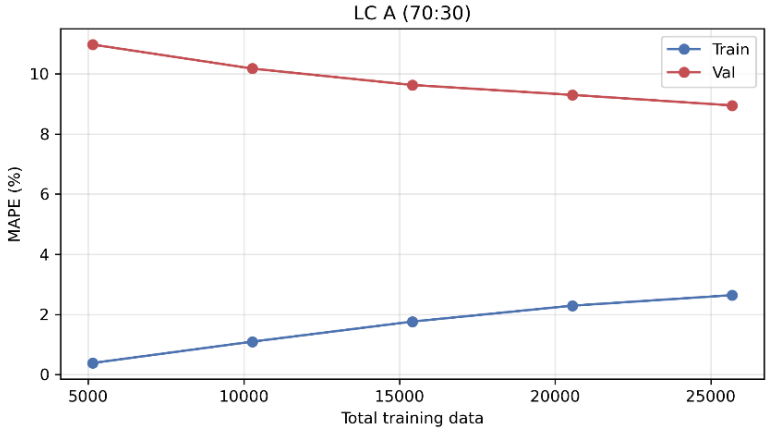
Parity plot menunjukkan penyebaran titik yang lebih lebar dibandingkan *split* 80:20 dan 90:10, terutama pada nilai ekstrem. Hal ini menunjukkan bahwa keterbatasan data latih mulai berdampak pada kemampuan model dalam melakukan prediksi secara presisi.

Tabel 4.15 Hasil Evaluasi Model A *Split* Data 70:30

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model A	0.9351	2.3875	0.9511	8.8529

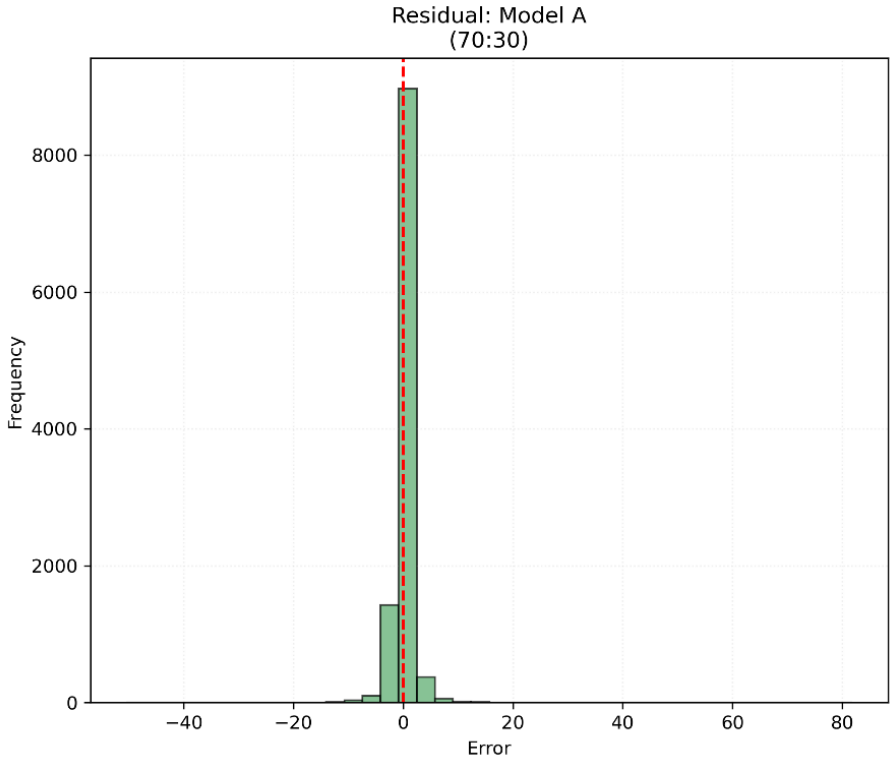
Model A menghasilkan MAE sebesar 0,9351 jam, RMSE sebesar 2,3875, R^2 sebesar 0,9511, dan MAPE sebesar 8.8529%. Nilai R^2 yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sekitar 95,11% variasi data target. Jika dibandingkan dengan skenario *split* sebelumnya, perbedaan performa yang

diperoleh relatif kecil sehingga menunjukkan bahwa model memiliki kemampuan generalisasi yang baik meskipun jumlah data latih berkurang menjadi 70% dari total dataset. Hasil ini mengindikasikan bahwa konfigurasi fitur lengkap pada Model A mampu menghasilkan prediksi yang stabil pada berbagai skenario pembagian data.



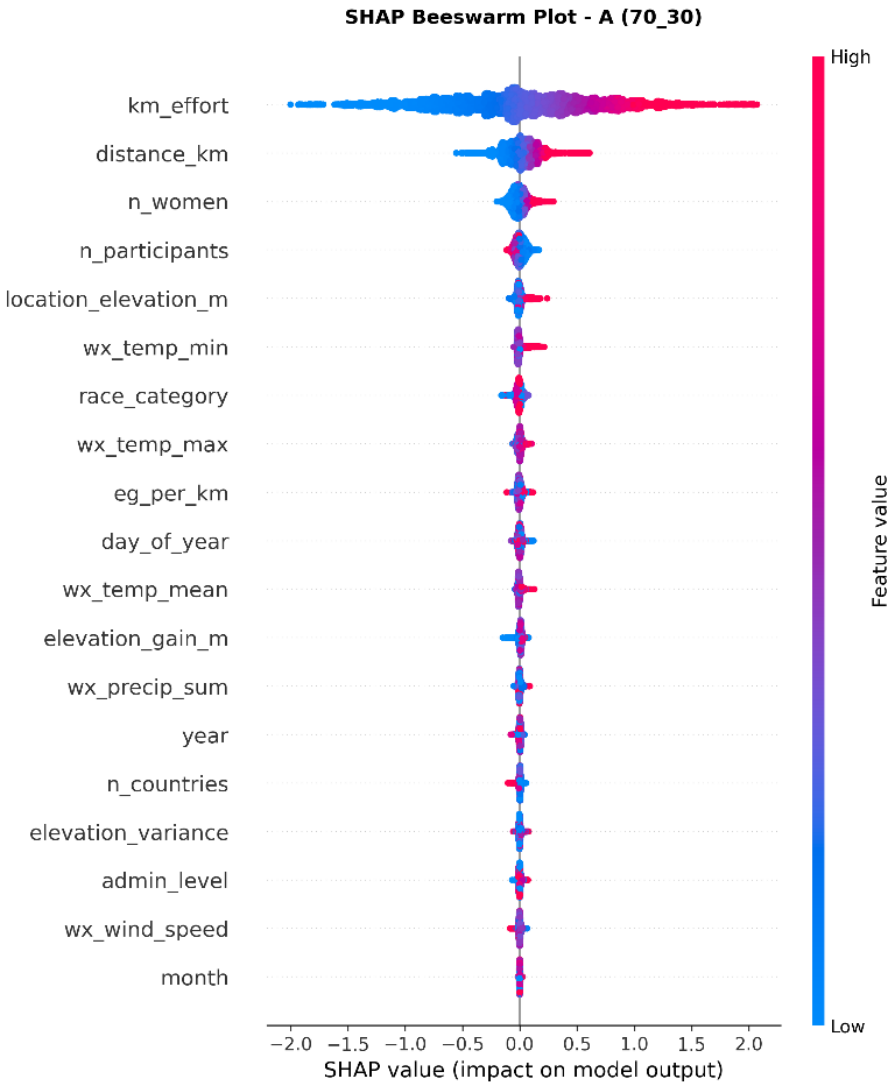
Gambar 4.44 *Learning curve* Model A *Split* 70:30

Kurva pembelajaran menunjukkan gap yang lebih besar antara training *error* dan *validation error* dibandingkan dua skenario sebelumnya. Hal ini mengindikasikan peningkatan *variance* akibat berkurangnya jumlah data latih, yang berdampak pada penurunan kemampuan generalisasi.



Gambar 4.45 Residual *Error* Model A *Split* 70:30

Distribusi residual menunjukkan penyebaran yang lebih luas dibandingkan *split* sebelumnya, yang mencerminkan peningkatan variansi *error* dan menurunnya stabilitas prediksi.



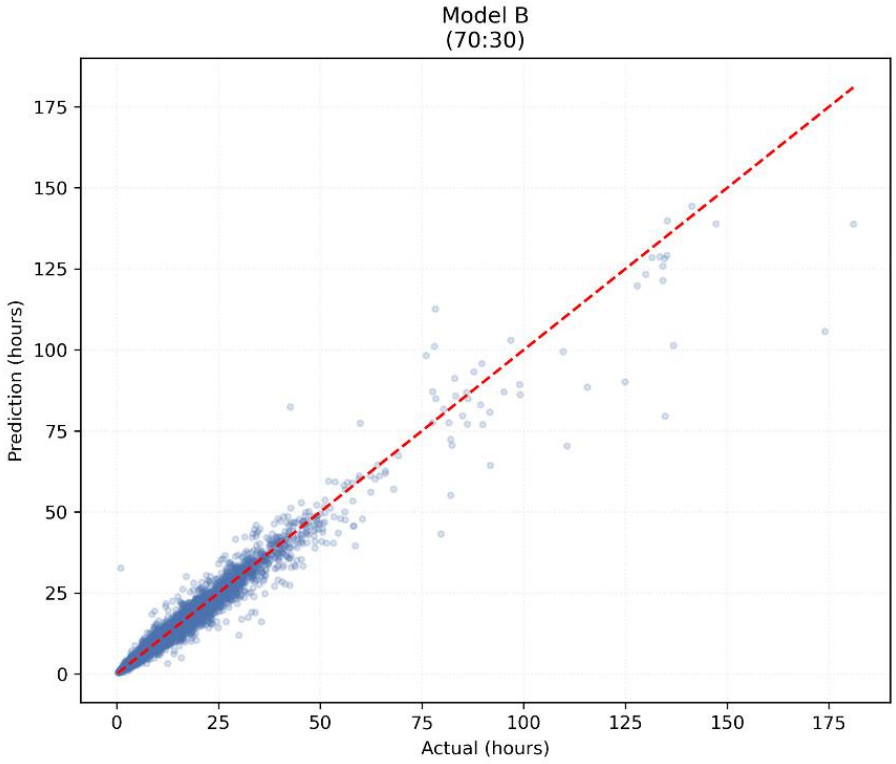
Gambar 4.46 SHAP Analisis Model A *Split* 70:30

Analisis SHAP tetap menunjukkan konsistensi kontribusi fitur, sehingga *distance_km* dan *elevation_gain_m* tetap mendominasi. Hal ini menunjukkan bahwa meskipun performa menurun, struktur hubungan antar fitur tetap stabil.

2. Model B

Model B merupakan skenario pemodelan yang menggunakan seluruh fitur karakteristik lintasan, namun menghilangkan fitur cuaca. Tujuan dari konfigurasi ini adalah untuk mengevaluasi sejauh mana kontribusi variabel lingkungan terhadap

performa model. Dengan membandingkan Model B terhadap Model A, dapat dianalisis dampak langsung dari absennya informasi cuaca terhadap akurasi prediksi.



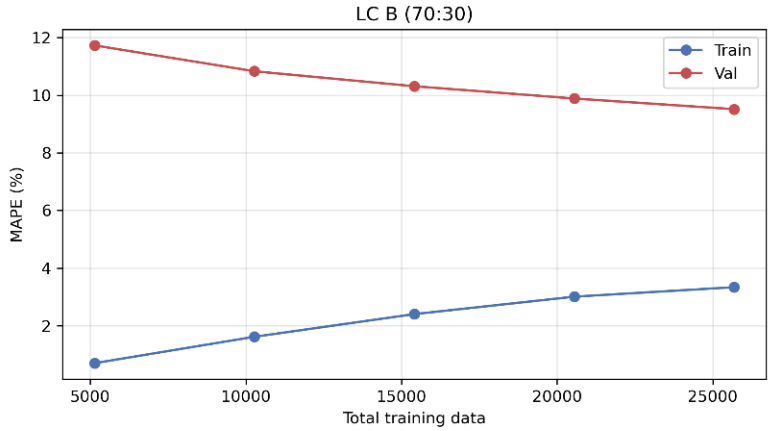
Gambar 4.47 *Parity plot* Model B *Split* 70:30

Pada visualisasi *parity plot*, titik-titik prediksi pada Model B masih mengikuti tren linear terhadap garis diagonal ($y = x$), namun dengan penyebaran yang lebih luas dibandingkan Model A. Penyimpangan ini terutama terlihat pada nilai waktu finis yang tinggi, yang mengindikasikan bahwa model mengalami kesulitan dalam menangkap variabilitas ekstrem tanpa dukungan fitur lingkungan.

Tabel 4.16 Hasil Evaluasi Model B *Split* Data 70:30

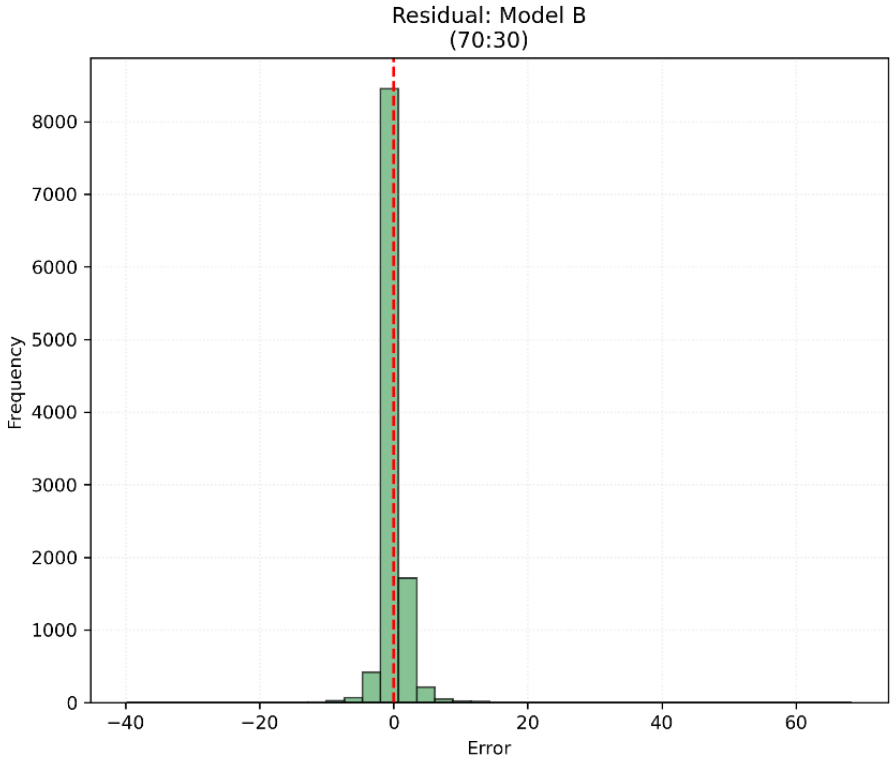
Model	MAE	RMSE	R2	MAPE (%)
Model B	0.9517	2.4027	0.9505	9.2586

Pada Tabel 4.16, Model B menghasilkan MAE sebesar 0,9517 jam, RMSE sebesar 2,4027, R^2 sebesar 0,9505, dan MAPE sebesar 9,2586%. Dibandingkan dengan Model A, terjadi peningkatan MAPE sebesar sekitar 0,41 poin persentase dan penurunan nilai R^2 sebesar 0,0006. Hasil ini menunjukkan bahwa penghapusan fitur cuaca menyebabkan berkurangnya informasi yang dapat dimanfaatkan model dalam memprediksi *mean finish time*. Meskipun demikian, nilai R^2 yang masih berada di atas 0,95 menunjukkan bahwa sebagian besar variasi target tetap dapat dijelaskan dengan baik menggunakan fitur selain cuaca.



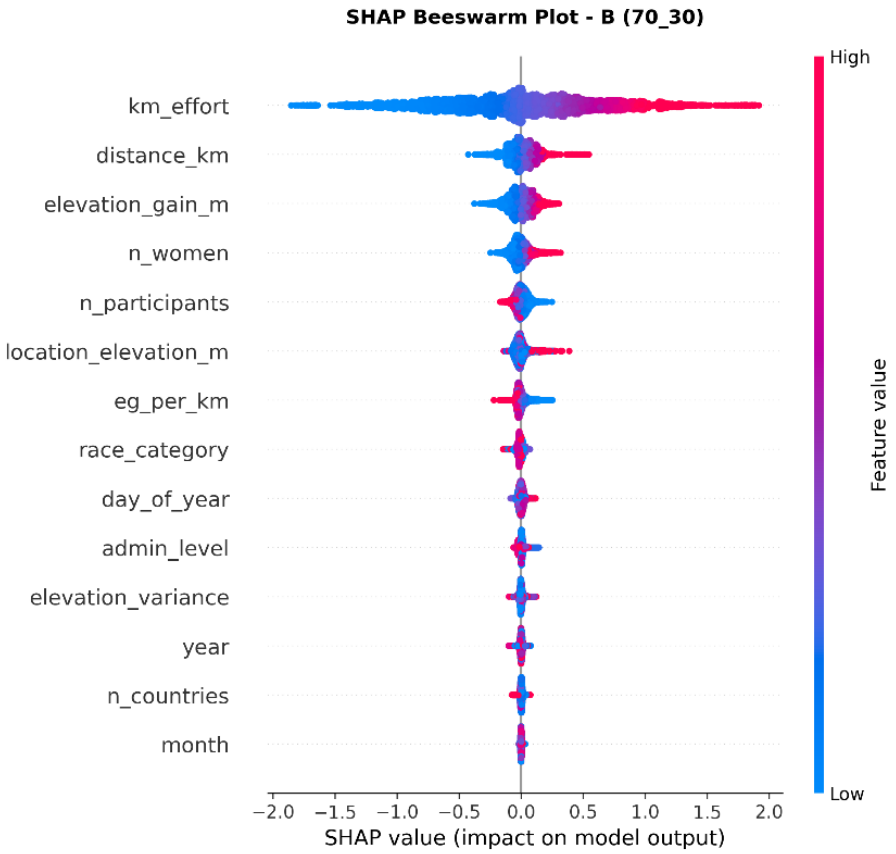
Gambar 4.48 *Learning curve* Model B *Split* 70:30

Kurva pembelajaran menunjukkan bahwa training *error* dan *validation error* tetap mengalami konvergensi, namun dengan gap yang lebih lebar dibandingkan Model A. Kondisi ini mengindikasikan adanya penurunan kemampuan generalisasi, yang disebabkan oleh kombinasi antara berkurangnya jumlah data latih (70%) dan hilangnya fitur penting. Dengan kata lain, model menjadi lebih sensitif terhadap keterbatasan informasi.



Gambar 4.49 Residual *Error* Model B *Split* 70:30

Distribusi residual menunjukkan pola yang masih terpusat di sekitar nol, namun dengan penyebaran yang lebih luas dibandingkan Model A. Hal ini mencerminkan peningkatan variansi *error*, terutama pada observasi dengan durasi lomba panjang, sehingga stabilitas prediksi menjadi menurun.



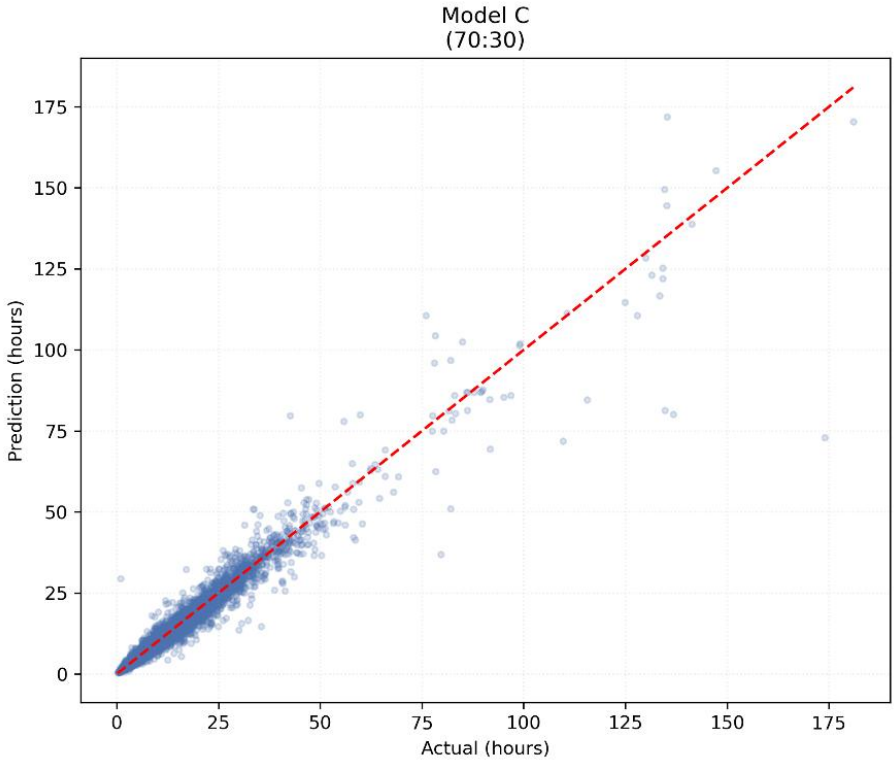
Gambar 4.50 SHAP Analisis Model B *Split* 70:30

Analisis SHAP menunjukkan bahwa fitur lintasan seperti *distance_km* dan *elevation_gain_m* tetap menjadi kontributor utama dalam model. Namun, absennya fitur cuaca menyebabkan berkurangnya kedalaman informasi yang digunakan dalam proses prediksi. Hal ini mengindikasikan bahwa variabel lingkungan sebelumnya berperan sebagai faktor penyesuaian penting dalam meningkatkan akurasi model.

3. Model C

Model C merupakan skenario pemodelan yang hanya menggunakan sepuluh fitur teratas berdasarkan tingkat kepentingan fitur (*Feature importance*) yang diperoleh dari proses sebelumnya. Pendekatan ini bertujuan untuk menguji apakah

model tetap dapat mempertahankan performa yang baik dengan jumlah fitur yang lebih sedikit, sekaligus meningkatkan efisiensi komputasi. Model ini merepresentasikan pendekatan reduksi dimensi berbasis seleksi fitur.



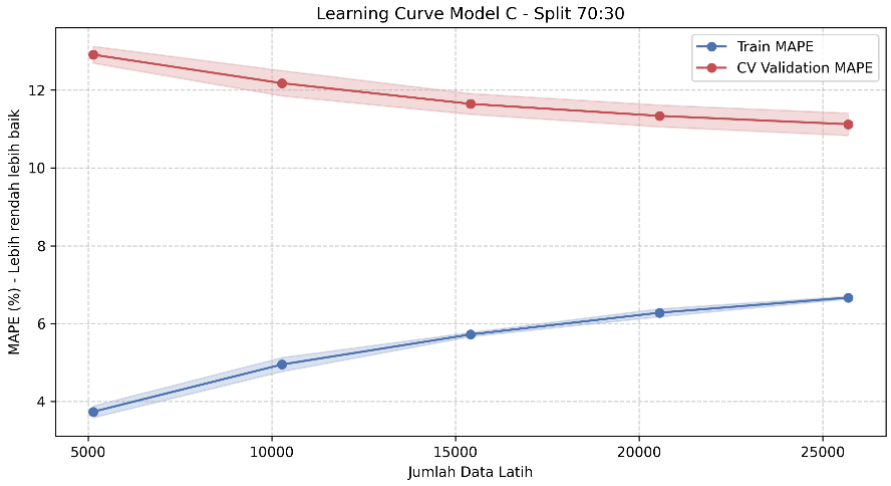
Gambar 4.51 *Parity plot* Model C *Split* 70:30

Pada *parity plot*, penyebaran titik prediksi terlihat semakin melebar dari garis diagonal dibandingkan Model B maupun Model A, terutama pada nilai ekstrem. Hal ini mengindikasikan bahwa model kehilangan kemampuan dalam menangkap hubungan *non-linear* yang kompleks akibat berkurangnya jumlah fitur.

Tabel 4.17 Hasil Evaluasi Model C *Split* Data 70:30

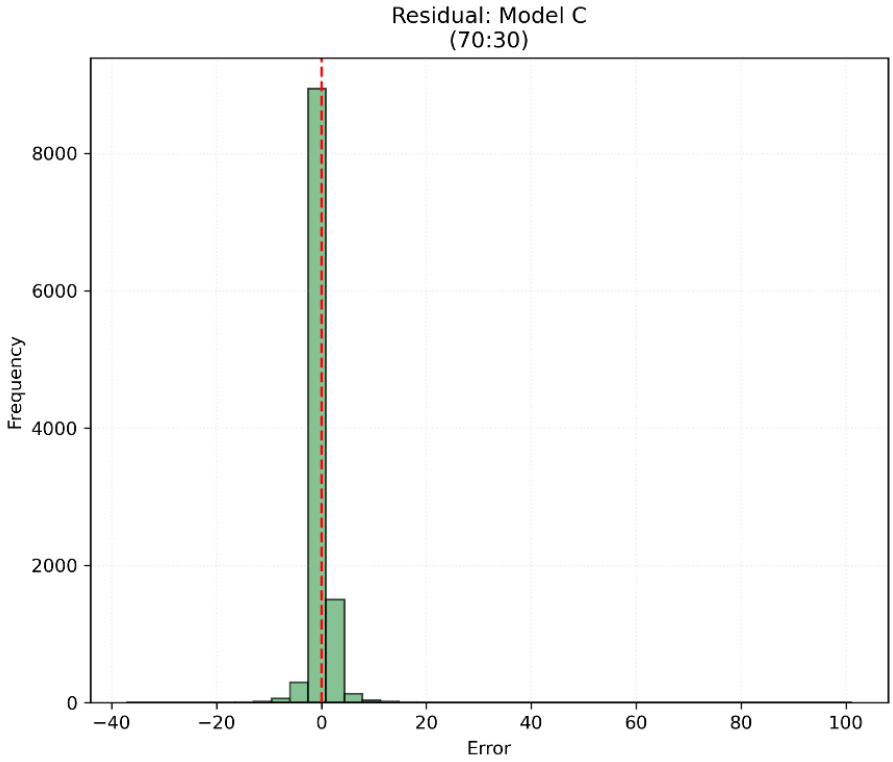
Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model C	0.9986	2.6052	0.9418	9.5464

Berdasarkan Tabel 4.17, Model C memperoleh MAE sebesar 0.9986 jam, RMSE sebesar 2.6052, R^2 sebesar 0,9418, dan MAPE sebesar 9.5464%. Dibandingkan dengan Model A, terjadi peningkatan MAPE sebesar sekitar 0,69 poin persentase serta penurunan nilai R^2 sebesar 0,0093. Hasil tersebut menunjukkan bahwa proses reduksi fitur menyebabkan sebagian informasi penting yang berkontribusi terhadap akurasi prediksi tidak lagi tersedia bagi model. Walaupun demikian, nilai MAPE yang masih berada pada kisaran 9% menunjukkan bahwa sepuluh fitur teratas hasil seleksi tetap mampu mempertahankan sebagian besar kemampuan prediktif model dengan jumlah fitur yang lebih ringkas.



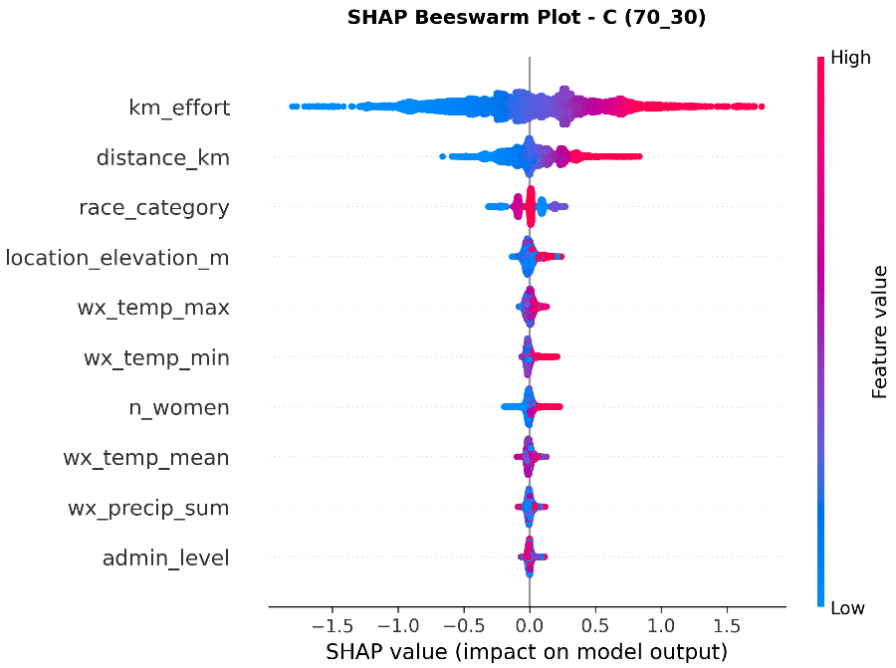
Gambar 4.52 *Learning curve* Model C *Split* 70:30

Kurva pembelajaran menunjukkan bahwa meskipun model tetap mencapai konvergensi, nilai *validation error* berada pada tingkat yang lebih tinggi dibandingkan Model A. Selain itu, gap antara training dan *validation error* cenderung meningkat, yang menandakan adanya penurunan kemampuan generalisasi.



Gambar 4.53 Residual *Error* Model C *Split* 70:30

Distribusi residual menunjukkan variansi yang lebih besar dibandingkan Model A, yang mencerminkan peningkatan ketidakstabilan dalam prediksi. Hal ini mengindikasikan bahwa model tidak mampu menjaga konsistensi performa pada seluruh rentang data.

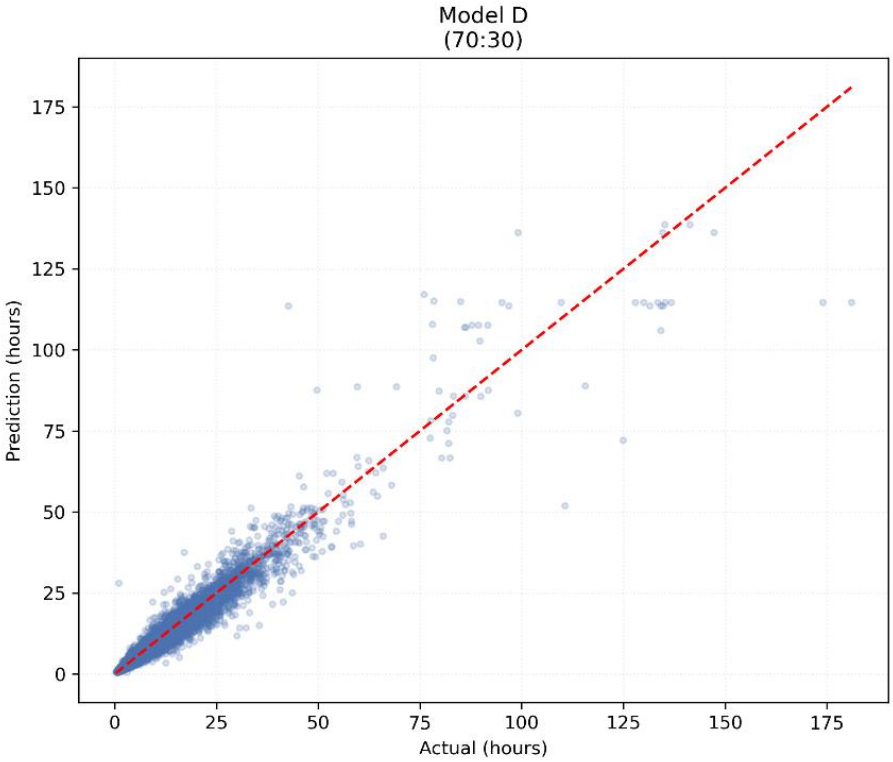


Gambar 4.54 SHAP Analisis Model C *Split* 70:30

Analisis SHAP memperlihatkan bahwa fitur utama seperti *km_effort*, *distance_km*, dan *elevation_gain_m* tetap mendominasi. Namun, dibandingkan Model A, distribusi kontribusi fitur menjadi lebih terkonsentrasi pada beberapa variabel saja. Hal ini menunjukkan bahwa model kehilangan kemampuan dalam menangkap interaksi kompleks antar fitur akibat reduksi dimensi.

4. Model D

Model D merupakan skenario pemodelan dengan menggunakan fitur minimal, yaitu hanya fitur-fitur dasar yang paling fundamental seperti *distance_km* dan *elevation_gain_m*. Model ini bertujuan untuk mengevaluasi batas bawah performa model ketika informasi yang digunakan sangat terbatas. Dengan demikian, Model D dapat digunakan sebagai pembanding untuk melihat sejauh mana kompleksitas fitur memengaruhi akurasi prediksi.



Gambar 4.55 *Parity plot* Model D *Split* 70:30

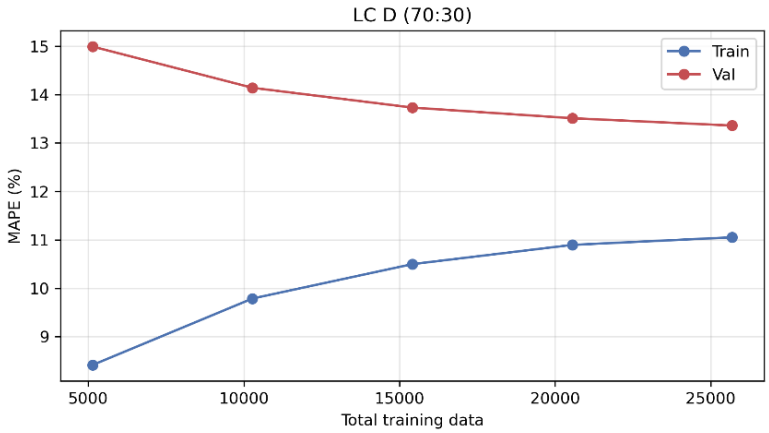
Pada *parity plot*, titik-titik prediksi menunjukkan deviasi yang jauh lebih besar dari garis diagonal, yang menandakan ketidakakuratan prediksi yang tinggi. Penyimpangan ini semakin jelas pada nilai waktu finis yang besar, yang menunjukkan ketidakmampuan model dalam menangkap pola kompleks.

Tabel 4.18 Hasil Evaluasi Model D *Split* Data 70:30

Model	**MAE**	**RMSE**	**R2**	**MAPE (%)**
Model D	1.3500	2.8098	0.9323	13.2360

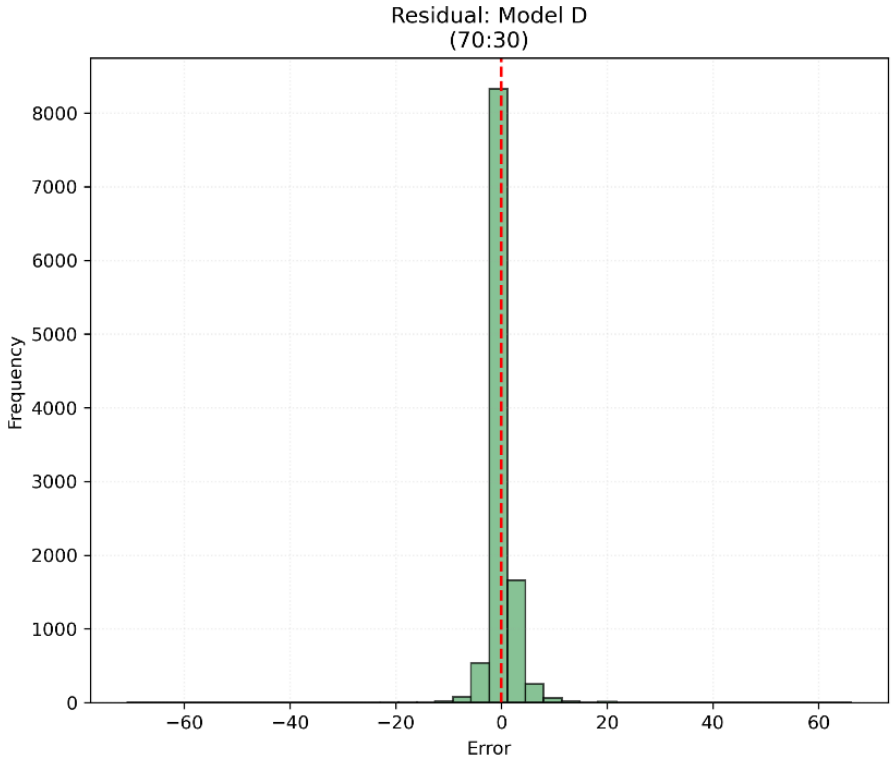
Model D menghasilkan MAE sebesar 1,35 jam, RMSE sebesar 2,8098, R^2 sebesar 0,9323, dan MAPE sebesar 13,2360%. Nilai ini merupakan performa terendah di antara seluruh model yang diuji pada skenario *split* 70:30. Jika dibandingkan dengan Model A, terjadi peningkatan MAPE sebesar sekitar 4,38

poin persentase serta penurunan nilai R^2 sebesar 0,02. Hasil tersebut menunjukkan bahwa penggunaan fitur minimal belum mampu menggambarkan kompleksitas hubungan antara karakteristik *event* dan *mean finish time* secara memadai. Oleh karena itu, fitur tambahan di luar jarak dan elevasi tetap diperlukan untuk menghasilkan prediksi yang lebih akurat dan konsisten.



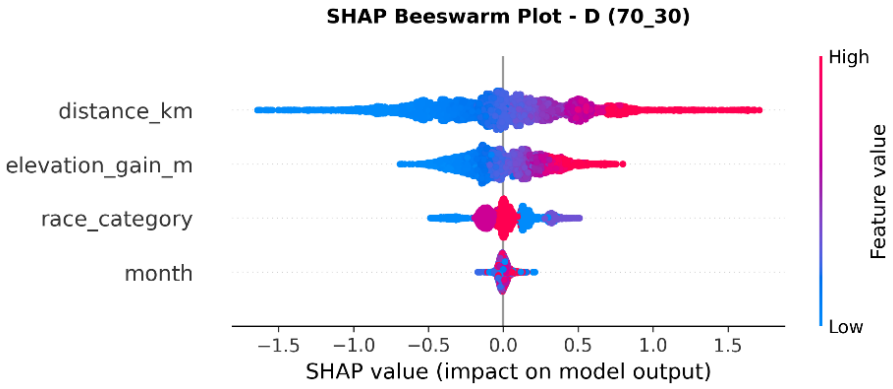
Gambar 4.56 *Learning curve* Model D *Split* 70:30

Kurva pembelajaran menunjukkan bahwa meskipun terjadi konvergensi, nilai *error* tetap tinggi dan cenderung stabil tanpa adanya penurunan signifikan. Hal ini mengindikasikan bahwa keterbatasan fitur menjadi bottleneck utama dalam peningkatan performa model, bahkan ketika data latih masih tersedia dalam jumlah yang cukup.



Gambar 4.57 Residual *Error* Model D *Split* 70:30

Distribusi residual menunjukkan penyebaran yang sangat luas dibandingkan Model A, yang menandakan tingginya variansi *error* dan rendahnya stabilitas prediksi. Hal ini mencerminkan bahwa model tidak mampu memberikan estimasi yang konsisten.



Gambar 4.58 SHAP Analisis Model D *Split* 70:30

Analisis SHAP mengonfirmasi bahwa model hanya bergantung pada sejumlah kecil fitur dasar, dengan kontribusi yang terbatas dan kurang variatif. Tidak adanya fitur tambahan menyebabkan model kehilangan konteks penting dalam data, sehingga tidak mampu merepresentasikan hubungan kompleks antar variabel secara memadai.

4.1.6 Hasil Kinerja Model

Pada tahap ini dilakukan evaluasi performa seluruh model berdasarkan empat metrik utama, yaitu MAE, RMSE, R^2, dan MAPE, pada tiga skenario pembagian data, yaitu 80:20, 70:30, dan 90:10. Hasil evaluasi secara keseluruhan disajikan pada Tabel 4.19.

Tabel 4.19 Hasil Kinerja Seluruh Model

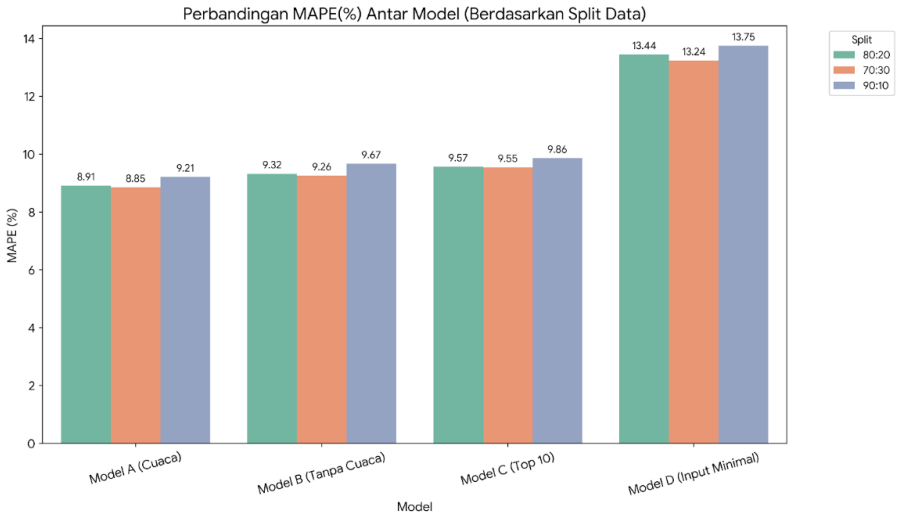
Split	**Model**	**MAE**	**RMSE**	**R2**	**MAPE (%)**
90:10	Model A	0.9047	2.2900	0.9570	9.2141
	Model B	0.9484	2.3025	0.9565	9.6732
	Model C	0.9776	2.4865	0.9493	9.8601
	Model D	1.4260	3.1259	0.9198	13.7510
80:20	Model A	0.9267	2.5732	0.9452	8.9117
	Model B	0.9427	2.3236	0.9553	9.3195
	Model C	0.9859	2.5779	0.9450	9.5688
	Model D	1.3564	2.8404	0.9333	13.4438
70:30	Model A	0.9351	2.3875	0.9511	8.8529
	Model B	0.9517	2.4027	0.9505	9.2586
	Model C	0.9986	2.6052	0.9418	9.5464
	Model D	1.3500	2.8098	0.9323	13.2360

Tabel 4.19 menyajikan ringkasan performa seluruh model pada seluruh skenario pengujian. Secara umum, Model A secara konsisten menghasilkan nilai

MAE, RMSE, dan MAPE terendah serta nilai R^2 tertinggi pada setiap skenario *split* data. Hasil ini menunjukkan bahwa kombinasi seluruh fitur yang tersedia mampu memberikan representasi paling lengkap terhadap faktor-faktor yang memengaruhi *mean_finish_time* pada *event Trail Running*. Dengan informasi yang lebih komprehensif, model dapat menangkap hubungan *non-linear* antar variabel secara lebih efektif sehingga menghasilkan prediksi yang lebih akurat.

Model B menempati posisi kedua pada seluruh skenario pengujian. Selisih performa antara Model B dan Model A relatif kecil, yang menunjukkan bahwa fitur cuaca memang memberikan kontribusi terhadap peningkatan akurasi, namun bukan merupakan faktor dominan dalam proses prediksi. Sebagian besar kemampuan prediktif model masih berasal dari fitur-fitur utama yang merepresentasikan karakteristik lintasan dan *event*.

Sementara itu, penurunan performa yang lebih besar terlihat pada Model C dan terutama Model D. Temuan ini menunjukkan bahwa pengurangan jumlah fitur menyebabkan hilangnya informasi yang penting bagi model dalam mempelajari pola data. Dengan demikian dapat disimpulkan bahwa akurasi prediksi *mean finish time* tidak hanya dipengaruhi oleh keberadaan fitur tertentu, tetapi juga oleh kelengkapan informasi yang tersedia dalam dataset.



Gambar 4.59 Perbandingan MAPE Seluruh Model

Gambar 4.59 memperlihatkan perbandingan nilai MAPE antar model pada seluruh skenario *split*. Terlihat bahwa Model A memiliki nilai MAPE terendah secara konsisten, sedangkan Model D memiliki nilai tertinggi. Perbedaan performa antar model terlihat jelas dan menunjukkan pola yang seragam pada setiap skenario *split*.

Secara keseluruhan, hasil ini menunjukkan bahwa kelengkapan fitur memiliki pengaruh yang kuat terhadap performa model. Model dengan fitur lengkap mampu menghasilkan prediksi yang lebih akurat dan stabil, sedangkan pengurangan fitur menyebabkan peningkatan *error* secara konsisten. Temuan ini menjadi dasar untuk analisis lebih lanjut pada Subbab 4.2 yang membahas secara khusus pengaruh variasi *split* data dan kelengkapan fitur terhadap performa model.

4.2 Pembahasan

4.2.1 Analisis *Robustness* Model terhadap Variasi *Split* Data

Analisis *Robustness* dilakukan untuk mengevaluasi konsistensi performa model terhadap variasi pembagian data latih dan data uji. Pada penelitian ini digunakan tiga skenario pembagian data, yaitu 90:10, 80:20, dan 70:30, dengan evaluasi berbasis *k-Fold Cross-Validation* (k=5) pada setiap scenario *training*.

Selanjutnya, analisis kuantitatif dilakukan dengan membandingkan nilai rata-rata performa model pada setiap skenario *split*, sebagaimana ditunjukkan pada Tabel 4.20.

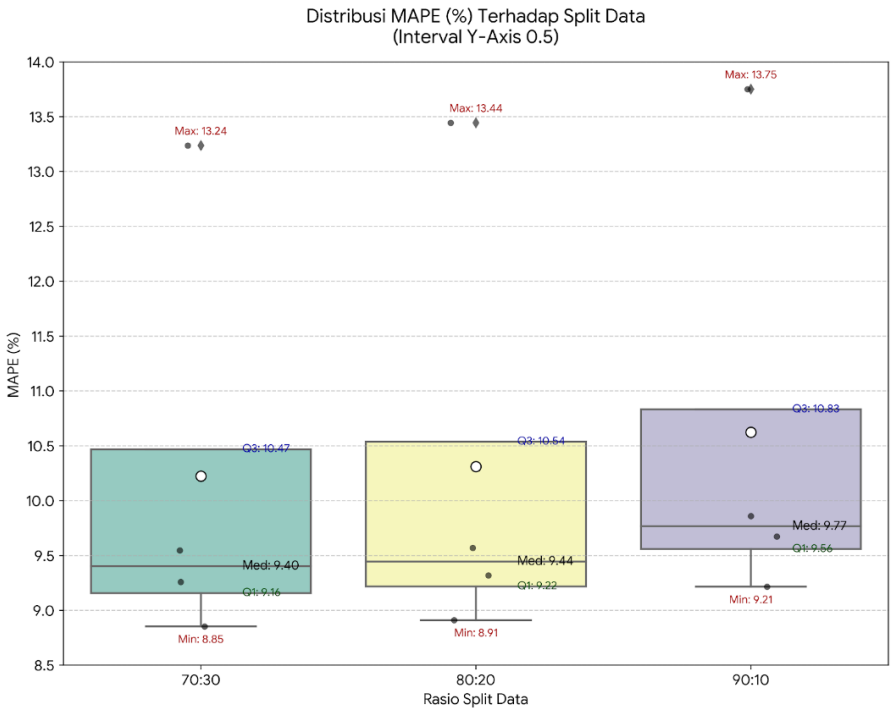
Tabel 4.20 Performa Setiap *Split*

Split	**MAE**	**RMSE**	**R2**	**MAPE (%)**
90:10	0.9086	2.5784	0.9403	8.5767
80:20	0.9329	2.7747	0.9315	8.7324
70:30	0.9409	2.6890	0.9328	8.9470

Berdasarkan Tabel 4.20, perbedaan performa antar skenario *split* relatif kecil pada seluruh metrik evaluasi. Nilai MAE berada pada rentang 0,9086 hingga 0,9409, sedangkan nilai RMSE berada pada rentang 2,5784 hingga 2,7747. Nilai R^2 juga relatif konsisten, yaitu antara 0,9315 hingga 0,9403. Pada metrik MAPE, nilai yang diperoleh berkisar antara 8,5767% hingga 8,9470%, dengan selisih maksimum hanya sekitar 0,37 poin persentase. Hasil ini menunjukkan bahwa perubahan proporsi data latih dan data uji tidak menyebabkan perubahan performa yang besar pada model.

Temuan ini mengindikasikan bahwa perubahan proporsi data latih dan data uji tidak memberikan dampak yang signifikan terhadap performa model. Untuk

memperkuat interpretasi tersebut, distribusi nilai *error* pada masing-masing skenario *split* divisualisasikan menggunakan boxplot.



Gambar 4.60 *Boxplot* Distribusi MAPE Terhadap *Split*

Seperti Tabel 4.20, pada Gambar 4.60 merupakan visualisasi *boxplot* yang menunjukkan bahwa distribusi performa memiliki pola yang relatif konsisten. Tidak terdapat perbedaan distribusi yang mencolok antar skenario, yang mengindikasikan bahwa perubahan proporsi data tidak memengaruhi distribusi *error* model secara signifikan.

Selain itu, sebaran nilai MAPE yang relatif sempit pada setiap skenario menunjukkan bahwa variasi performa antar fold cukup rendah. Hal ini mengindikasikan bahwa model memiliki stabilitas yang baik terhadap variasi data.

Untuk memastikan bahwa perbedaan performa antar skenario *split* tidak terjadi secara kebetulan, dilakukan uji statistik menggunakan *paired t-test* berbasis

hasil evaluasi pada setiap *fold Cross-Validation*. Hasil pengujian ditunjukkan pada Tabel 4.20.

Tabel 4.21 Uji *T-Test* Pengaruh *Split* Data

Skenario	Metrik Evaluasi	p-value	α	Keterangan
90:10 vs 80:20	MAE	0.363	0.05	Tidak Signifikan
	RMSE	0.816	0.05	Tidak Signifikan
	R^2	0.801	0.05	Tidak Signifikan
	MAPE	0.214	0.05	Tidak Signifikan
80:20 vs 70:30	MAE	0.808	0.05	Tidak Signifikan
	RMSE	0.926	0.05	Tidak Signifikan
	R^2	0.974	0.05	Tidak Signifikan
	MAPE	0.010	0.05	Signifikan
90:10 vs 70:30	MAE	0.170	0.05	Tidak Signifikan
	RMSE	0.903	0.05	Tidak Signifikan
	R^2	0.852	0.05	Tidak Signifikan
	MAPE	0.031	0.05	Signifikan

Hasil uji *paired t-test* pada Tabel 4.21 menunjukkan bahwa seluruh perbandingan pada metrik MAE, RMSE, dan R^2 menghasilkan nilai *p-value* yang lebih besar dari 0,05. Hal ini menunjukkan bahwa perubahan proporsi data latih dan data uji tidak memberikan pengaruh yang signifikan terhadap ketiga metrik tersebut. Namun, pada metrik MAPE ditemukan perbedaan yang signifikan antara skenario *split* 80:20 dan 70:30 (p = 0,010) serta antara *split* 90:10 dan 70:30 (p = 0,031). Meskipun demikian, selisih nilai MAPE yang terjadi relatif kecil, yaitu kurang dari 0,4 poin persentase, sehingga secara praktis performa model tetap dapat dianggap stabil terhadap variasi pembagian data.

Karena hasil evaluasi dan pengujian statistik pada model menunjukkan tingkat *Robustness* yang baik. Oleh karena itu, analisis selanjutnya difokuskan pada pengaruh kelengkapan fitur terhadap performa model. Skenario *split* 90:10 dipilih sebagai skenario representatif karena menghasilkan nilai R^2 tertinggi dan merupakan konfigurasi dengan proporsi data latih terbesar, sehingga memungkinkan model memanfaatkan informasi pelatihan secara lebih optimal.

4.2.2 Analisis Pengaruh Fitur terhadap Performa Model

Berdasarkan hasil analisis *Robustness* pada Subbab 4.2.1, diketahui bahwa variasi pembagian data tidak memberikan pengaruh signifikan terhadap performa model. Oleh karena itu, analisis pada subbab ini difokuskan pada pengaruh kelengkapan fitur terhadap performa model dengan menggunakan satu skenario representatif, yaitu *split* 90:10.

Analisis Kinerja Model (*Split* 90:10)

Tabel 4.22 Tabel Analisis *Split* 90:10

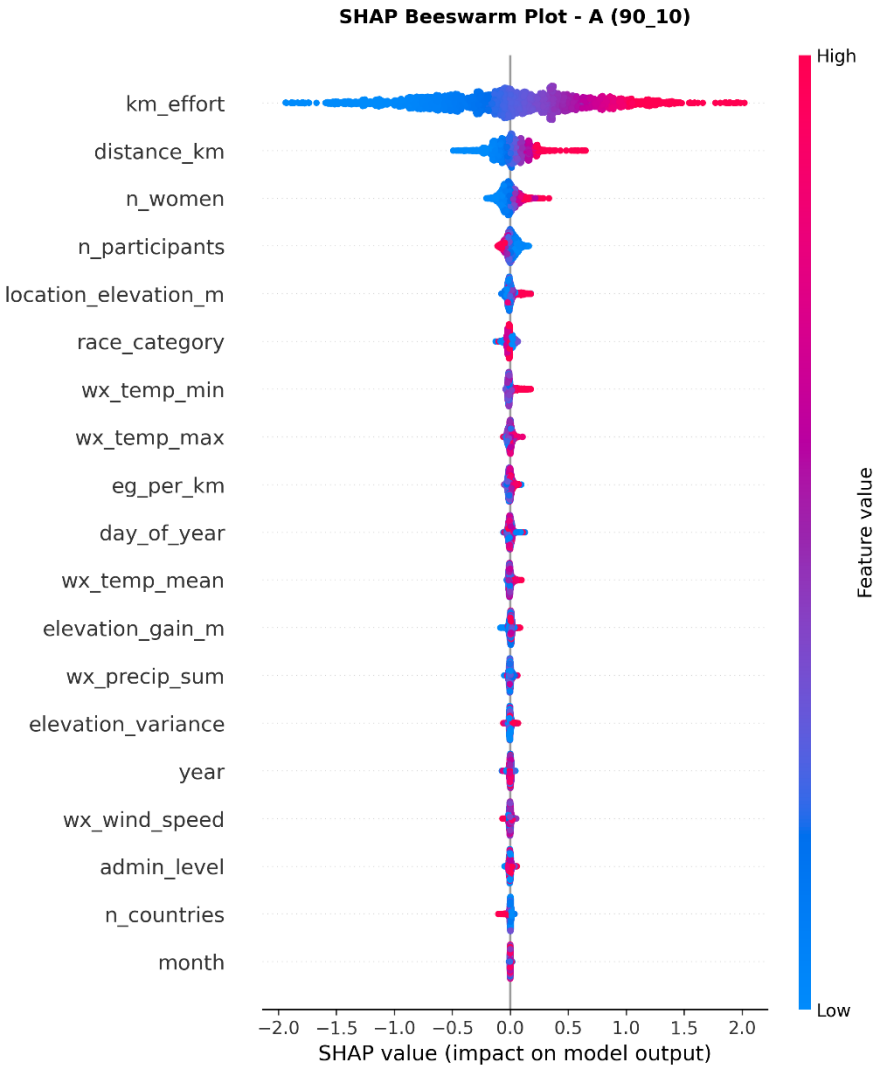
Model / Statistik	MAE	RMSE	R2	MAPE (%)
Model A	0.9047	2.2900	0.9570	9.2141
Model B	0.9484	2.3025	0.9565	9.6732
Model C	0.9776	2.4865	0.9493	9.8601
Model D	1.4260	3.1259	0.9198	13.7510

Berdasarkan Tabel 4.22, pengurangan jumlah fitur menyebabkan peningkatan *error* pada seluruh metrik evaluasi. Jika dibandingkan dengan Model A sebagai *Baseline*, Model B mengalami peningkatan MAPE sebesar sekitar 4,98%, Model C sebesar 7,01%, dan Model D sebesar 49,24%. Pola yang sama juga terlihat pada metrik MAE dan RMSE yang meningkat secara bertahap seiring

berkurangnya jumlah fitur yang digunakan. Sementara itu, nilai R^2 mengalami penurunan dari 0,9570 pada Model A menjadi 0,9565 pada Model B, 0,9493 pada Model C, dan 0,9198 pada Model D. Hasil ini menunjukkan bahwa kelengkapan fitur berperan penting dalam mempertahankan kemampuan model untuk menjelaskan variasi *mean finish time*.

Meskipun seluruh model masih menunjukkan performa yang cukup baik, hasil evaluasi menunjukkan bahwa pengurangan fitur secara bertahap selalu diikuti oleh peningkatan *error* prediksi. Menariknya, selisih performa antara Model A dan Model B relatif kecil, yang mengindikasikan bahwa fitur cuaca berfungsi sebagai fitur pendukung yang meningkatkan akurasi model secara moderat. Sebaliknya, penurunan performa yang lebih besar pada Model C dan terutama Model D menunjukkan bahwa sebagian besar informasi penting untuk prediksi berasal dari kombinasi berbagai fitur karakteristik lintasan dan *event* yang digunakan pada Model A.

Analisis Interpretabilitas Model (SHAP)



Gambar 4.61 SHAP Model A

Gambar 4.61 menunjukkan hasil analisis SHAP pada model terbaik (Model A). Berdasarkan visualisasi tersebut, fitur *km_effort* memiliki kontribusi paling dominan terhadap prediksi, diikuti oleh *elevation_gain_m* dan *distance_km*. Hal ini

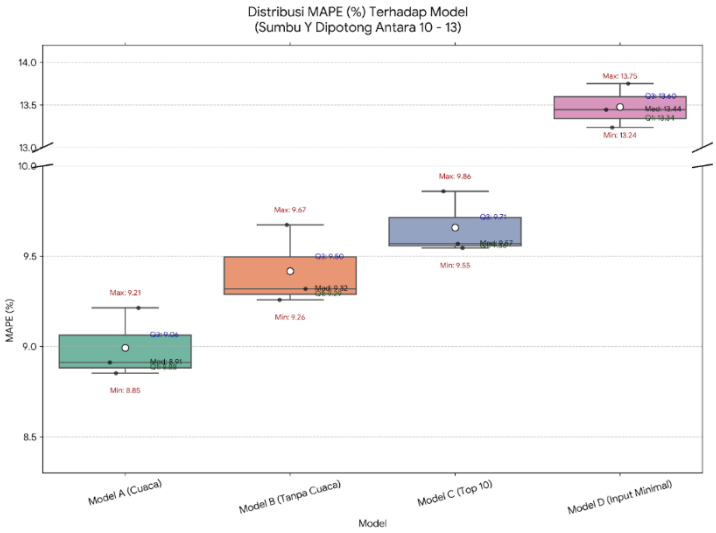
menunjukkan bahwa karakteristik lintasan merupakan faktor utama dalam menentukan durasi lomba.

Selain itu, fitur lingkungan seperti *wx_temp_min* dan *wx_temp_max* juga menunjukkan kontribusi yang signifikan. Nilai SHAP yang positif pada nilai fitur yang tinggi (ditunjukkan dengan warna merah) mengindikasikan bahwa suhu yang lebih tinggi cenderung meningkatkan waktu finis.

Sebaliknya, fitur seperti *wx_wind_speed*, *wx_precip_sum*, dan *day_of_year* memiliki kontribusi relatif kecil, yang terlihat dari distribusi nilai SHAP yang terpusat di sekitar nol. Hal ini menunjukkan bahwa fitur tersebut memiliki pengaruh terbatas terhadap prediksi model. Dengan demikian, analisis SHAP mengonfirmasi bahwa:

1. Fitur lintasan merupakan faktor dominan dalam prediksi.
2. Fitur cuaca berperan sebagai variabel pendukung yang meningkatkan presisi model.

Uji Statistik Antar Model (*Paired t-test*)



Gambar 4.62 *Boxplot* Distribusi MAPE Terhadap Model

Gambar 4.62 menunjukkan bahwa Model A memiliki distribusi *error* paling sempit dengan median terendah, yang mengindikasikan performa paling stabil dan akurat. Model B dan Model C menunjukkan peningkatan median *error* dengan distribusi yang sedikit lebih lebar, sedangkan Model D memiliki median tertinggi yang menunjukkan peningkatan *error* secara konsisten.

Perbedaan distribusi ini memberikan indikasi awal bahwa pengurangan fitur berdampak terhadap performa model. Untuk memastikan apakah perbedaan tersebut signifikan secara statistik, dilakukan uji *paired t-test* berbasis hasil evaluasi pada setiap *fold Cross-Validation.*

Tabel 4.23 Uji T-Test Model Lain Terhadap Model A

Perbandingan	**Metrik**	**Mean Model A**	**Mean Pembanding**	**p-value**	**Signifikansi**
Model A vs B	MAE	0.9274	0.9715	<0.001	Signifikan
	RMSE	2.6807	2.6967	0.504	Tidak Signifikan
	R^2	0.9349	0.9337	0.327	Tidak Signifikan
	MAPE	8.7521	9.2943	<0.001	Signifikan

Perbandingan	Metrik	Mean Model A	Mean Pembanding	p-value	Signifikansi
Model A vs C	MAE	0.9274	0.9942	<0.001	Signifikan
	RMSE	2.6807	2.8316	0.038	Signifikan
	R^2	0.9349	0.9283	0.038	Signifikan
	MAPE	8.7521	9.4987	<0.001	Signifikan
Model A vs D	MAE	0.9274	1.3625	<0.001	Signifikan
	RMSE	2.6807	3.2764	<0.001	Signifikan
	R^2	0.9349	0.9059	<0.001	Signifikan
	MAPE	8.7521	13.1715	<0.001	Signifikan

Hasil uji statistik pada Tabel 4.23 menunjukkan bahwa perbedaan antara Model A dan Model B hanya signifikan pada metrik MAE dan MAPE, sedangkan RMSE dan R^2 tidak menunjukkan perbedaan yang signifikan ($p > 0,05$). Temuan ini menunjukkan bahwa penghapusan fitur cuaca memang menyebabkan penurunan akurasi prediksi, tetapi dampaknya relatif terbatas terhadap kemampuan model secara keseluruhan.

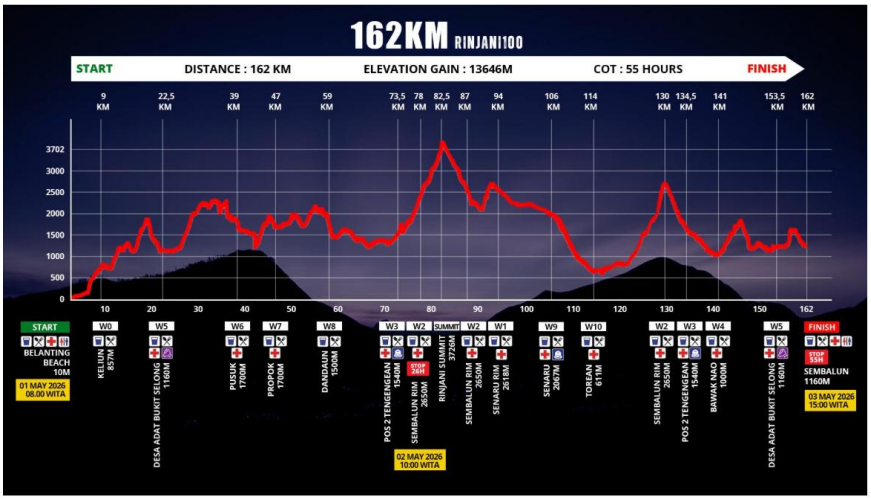
Pada perbandingan Model A dan Model C, seluruh metrik menunjukkan perbedaan yang signifikan dengan nilai *p-value* di bawah 0,05. Hasil ini menunjukkan bahwa reduksi fitur hingga hanya menggunakan sepuluh fitur teratas mulai mengurangi kemampuan model dalam menangkap hubungan kompleks antar variabel. Dengan demikian, sebagian fitur yang dieliminasi masih memiliki kontribusi yang penting terhadap performa prediksi.

Sementara itu, perbandingan antara Model A dan Model D menunjukkan perbedaan signifikan pada seluruh metrik evaluasi dengan nilai *p-value* yang sangat kecil (<0,001). Hasil ini mengindikasikan bahwa penggunaan fitur minimal menyebabkan hilangnya informasi penting yang diperlukan model untuk

menjelaskan variasi *mean finish time*. Oleh karena itu, meskipun fitur dasar seperti jarak dan elevasi memiliki pengaruh yang besar, kombinasi fitur yang lebih lengkap tetap diperlukan untuk menghasilkan performa prediksi yang optimal.

4.2.3 Implementasi Sistem Dan Manfaat Praktis

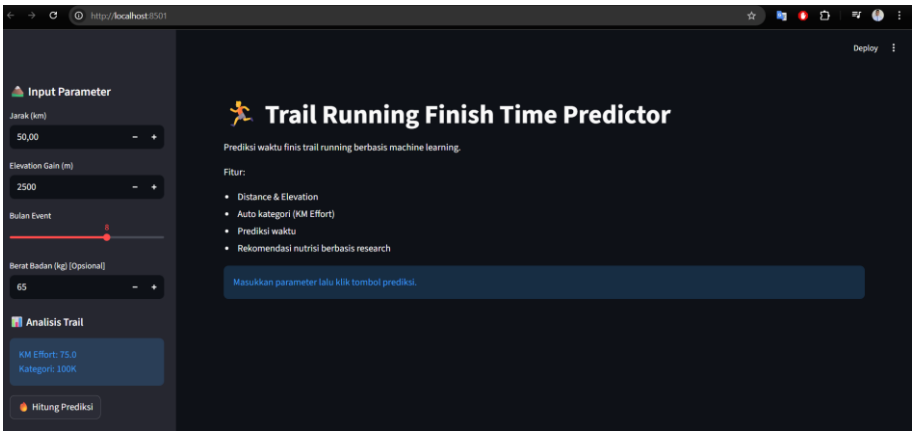
Berdasarkan hasil evaluasi pada subbab sebelumnya, Model A memiliki performa terbaik dari segi akurasi. Namun, pada tahap implementasi, digunakan Model D karena lebih sederhana dan hanya membutuhkan *input* fitur yang mudah diperoleh oleh pengguna. Pemilihan ini bertujuan agar sistem lebih praktis dan dapat digunakan dalam kondisi nyata.



Gambar 4.63 Contoh Profil Lintasan Trail

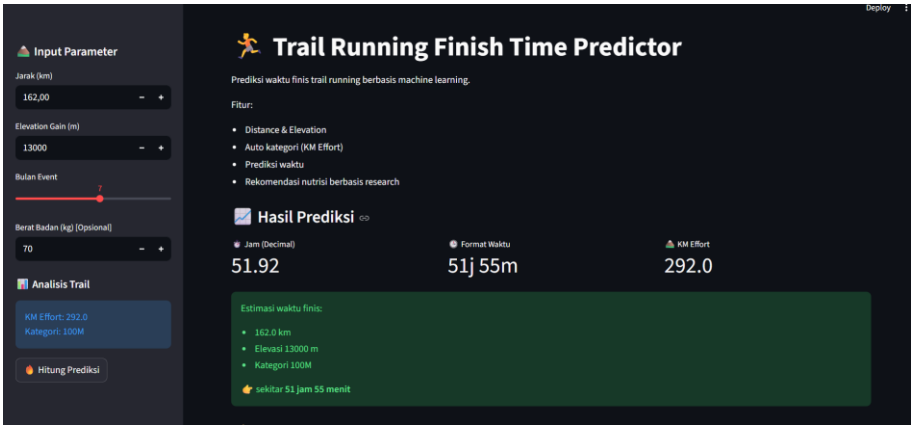
Gambar 4.63 menunjukkan profil elevasi lintasan sepanjang 162 km dengan total elevasi sekitar 13.646 meter. Lintasan memiliki variasi tanjakan dan turunan yang signifikan, termasuk segmen menuju puncak (*summit*) sebagai titik tertinggi.

Kondisi ini menunjukkan bahwa lintasan memiliki tingkat kesulitan tinggi, yang direpresentasikan dalam model melalui fitur *distance_km* dan *elevation_gain_m*.



Gambar 4.64 Antarmuka *Input* Sistem

Gambar 4.64 menampilkan antarmuka sistem prediksi yang dirancang sederhana dan mudah digunakan. Pengguna hanya perlu memasukkan parameter utama seperti jarak, elevasi, bulan *event*, dan berat badan (opsional). Sistem kemudian secara otomatis menghitung *km_effort* dan kategori lintasan, sehingga mempermudah proses prediksi.



Gambar 4.65 Hasil Prediksi Sistem

Gambar 4.65 menunjukkan hasil prediksi untuk lintasan 162 km dengan elevasi 13.000 meter. Sistem menghasilkan estimasi waktu finis sekitar 51,92 jam (51 jam 55 menit) dengan nilai *km_effort* sebesar 292,0.

Nilai tersebut menunjukkan bahwa lintasan termasuk kategori 100 Miles dengan tingkat kesulitan tinggi. Estimasi waktu juga masih berada dalam batas *cut-off time* (COT) sekitar 55 jam, sehingga dapat dianggap realistis.

Sistem ini membantu peserta dalam memperkirakan waktu tempuh secara sederhana berdasarkan jarak dan elevasi lintasan. Selain itu, indikator *km_effort* membantu memahami tingkat kesulitan lintasan sehingga dapat digunakan untuk perencanaan strategi lomba. Bagi penyelenggara, sistem ini dapat digunakan sebagai referensi dalam menilai tingkat kesulitan lintasan dan memberikan estimasi waktu kepada peserta.

Dalam penelitian ini, proses pengembangan model prediksi menunjukkan bahwa performa model sangat dipengaruhi oleh kelengkapan fitur yang digunakan. Berdasarkan hasil pengujian, pengurangan fitur secara bertahap memberikan dampak yang signifikan terhadap peningkatan *error* prediksi, khususnya pada metrik akurasi seperti MAE dan MAPE. Model dengan fitur lengkap (Model A) terbukti mampu memberikan keseimbangan terbaik antara akurasi dan stabilitas, sedangkan pengurangan fitur, baik secara parsial maupun ekstrem, menyebabkan penurunan performa yang sistematis. Hal ini mengindikasikan bahwa tidak terdapat fitur yang sepenuhnya tidak berkontribusi, melainkan setiap fitur memiliki peran dalam membangun model yang optimal secara kolektif.

Prinsip tersebut sejalan dengan firman Allah SWT dalam Al-Qur'an pada QS. Al-Qamar ayat 49:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ

"Sesungguhnya Kami menciptakan segala sesuatu menurut ukuran." (*QS. Al-Qamar: 49)*

Menurut tafsir Kementerian Agama Republik Indonesia, ayat ini menjelaskan bahwa seluruh ciptaan Allah SWT berlangsung berdasarkan ketentuan, ukuran, dan keteraturan yang telah ditetapkan-Nya. Tidak ada sesuatu pun yang tercipta secara sia-sia atau tanpa fungsi tertentu (*Qur'an Kemenag*, n.d.). Dalam konteks penelitian ini, konsep tersebut tercermin pada penggunaan fitur dalam model prediksi, sehingga setiap variabel memiliki kontribusi tertentu dalam membentuk performa model. Hasil penelitian menunjukkan bahwa pengurangan fitur secara bertahap diikuti oleh penurunan akurasi prediksi, yang mengindikasikan bahwa setiap fitur menyimpan informasi yang berperan dalam menjelaskan variasi *mean finish time*. Dengan demikian, model yang optimal diperoleh melalui keseimbangan dan keterpaduan antar fitur yang digunakan.

Disisi lain pengembangan model prediksi dalam penelitian ini juga merupakan bagian dari upaya pemanfaatan ilmu pengetahuan untuk membantu pengambilan keputusan yang lebih akurat. Hal ini selaras dengan hadis Rasulullah yang diriwayatkan dalam Sahih al-Bukhari:

مَا أَنْزَلَ اللَّهُ دَاءً إِلَّا أَنْزَلَ لَهُ دَوَاءً

"Tidaklah Allah menurunkan suatu penyakit, melainkan Dia juga menurunkan obatnya." (HR. Bukhari)

Menurut penjelasan para ulama dalam syarah hadis, sabda Rasulullah SAW tersebut menunjukkan bahwa Allah SWT tidak menurunkan suatu permasalahan tanpa menyediakan jalan keluar atau solusi yang dapat ditemukan melalui usaha manusia. Dalam *Fath al-Bari*, dijelaskan bahwa hadis ini mendorong umat Islam untuk melakukan ikhtiar, penelitian, dan pencarian ilmu dalam menemukan manfaat yang tersembunyi di balik ciptaan Allah SWT (Amiruddin, 2009). Oleh karena itu, manusia tidak diperintahkan untuk bersikap pasif, melainkan aktif melakukan pengamatan, pembelajaran, dan pengembangan pengetahuan guna memperoleh solusi atas berbagai permasalahan yang dihadapi.

Dalam konteks penelitian ini, pengembangan model prediksi menggunakan algoritma XGBoost merupakan salah satu bentuk ikhtiar ilmiah untuk menemukan pola dan informasi yang terkandung dalam data. Proses pengumpulan data, pemilihan fitur, optimasi *hyperparameter*, serta evaluasi model dilakukan secara sistematis untuk memperoleh model yang lebih akurat dan *robust*.

Jadi hasil penelitian ini tidak hanya menunjukkan bahwa kelengkapan fitur merupakan faktor utama dalam meningkatkan performa model, tetapi juga merefleksikan nilai keseimbangan dan keteraturan dalam setiap komponen yang digunakan. Proses optimasi model yang dilakukan bukanlah hasil kebetulan, melainkan hasil dari pemanfaatan struktur data secara tepat dan proporsional, yang pada akhirnya memberikan manfaat dalam pengembangan sistem prediksi yang lebih akurat dan dapat diandalkan.

BAB V

KESIMPULAN DAN SARAN

Kesimpulan dan saran harus dinyatakan secara terpisah, yaitu pada sub bab tersendiri di bab terakhir.

5.1 Kesimpulan

Berdasarkan hasil implementasi, pengujian, dan pembahasan yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Pengembangan model prediksi *mean finish time* menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) berhasil menghasilkan performa yang baik dan stabil pada berbagai skenario pembagian data. Model A yang menggunakan seluruh fitur menghasilkan performa terbaik secara keseluruhan dengan MAPE terendah 8,8529% (*split* 70:30), MAE terendah 0,9047 jam dan R^2 tertinggi 0,9570 (*split* 90:10). Selain itu, hasil evaluasi *Robustness* menunjukkan bahwa model memiliki tingkat stabilitas yang baik terhadap variasi pembagian data, ditunjukkan oleh nilai performa yang relatif konsisten pada skenario *split* 90:10, 80:20, dan 70:30.
2. Karakteristik lintasan merupakan faktor utama yang memengaruhi prediksi *mean finish time*, sedangkan fitur cuaca berperan sebagai variabel pendukung yang meningkatkan akurasi model. Hasil analisis SHAP menunjukkan bahwa fitur km_*effort*, *elevation_gain_m*, dan *distance_km* menjadi kontributor paling dominan dalam proses prediksi. Sementara itu, hasil perbandingan antar model menunjukkan bahwa penghapusan fitur cuaca pada Model B hanya

menyebabkan peningkatan *error* yang relatif kecil dibandingkan Model A. Temuan ini menunjukkan bahwa akurasi prediksi terbaik diperoleh melalui kombinasi fitur yang lengkap, namun sebagian besar kemampuan prediktif model tetap berasal dari karakteristik lintasan dan *event*.

5.2 Saran

Untuk pengembangan penelitian selanjutnya, beberapa hal yang dapat ditingkatkan (di-*improve*) berdasarkan hasil pembahasan adalah:

1. Integrasi variabel teknikalitas lintasan. Penelitian selanjutnya dapat menambahkan variabel yang merepresentasikan tingkat teknikalitas medan, seperti jenis permukaan lintasan, tingkat kesulitan jalur, keberadaan bebatuan, akar pohon, lumpur, maupun tingkat keterjalaman jalur. Variabel tersebut berpotensi meningkatkan kemampuan model dalam menangkap karakteristik *event Trail Running* yang belum sepenuhnya terwakili oleh jarak dan elevasi.
2. Pengembangan model untuk *event* ekstrem. Diperlukan kajian khusus terhadap *event ultra-trail* dengan durasi sangat panjang karena kelompok *event* tersebut masih menunjukkan variasi *error* yang lebih tinggi dibandingkan *event* dengan durasi yang lebih pendek. Pendekatan segmentasi data atau pengembangan model khusus untuk kategori *ultra-distance* dapat menjadi alternatif yang layak dieksplorasi.
3. Pengayaan sumber data cuaca. Penelitian selanjutnya dapat memanfaatkan data cuaca dengan resolusi temporal yang lebih tinggi, seperti data per jam (*hourly weather data),* sehingga perubahan kondisi cuaca selama perlombaan dapat direpresentasikan dengan lebih baik dibandingkan penggunaan data harian.

4. Evaluasi potensi *group leakage* pada data historis *event*. Dataset yang digunakan dalam penelitian ini mengandung beberapa *event* yang muncul berulang pada tahun yang berbeda. Kondisi tersebut berpotensi menyebabkan kemiripan karakteristik antara data latih dan data uji sehingga dapat meningkatkan performa model secara tidak langsung (*group leakage).* Oleh karena itu, penelitian selanjutnya disarankan menggunakan pendekatan validasi berbasis grup, seperti *Group K-Fold* atau *Leave-One-Event-Out Validation*, agar kemampuan generalisasi model terhadap *event* yang benar-benar baru dapat dievaluasi secara lebih ketat dan realistis.

DAFTAR PUSTAKA

Amiruddin;, I. H. A. A. (2009). *Fathul Baari : Penjelasan Kitab Shahih Al Bukhari (Buku 30)*. //libcat.uin-malang.ac.id

Anastasiadou, M., dos Santos, V. D., & Dias, M. S. (2025). An explainable deep *learning* model for energy performance classification and retrofitting recommendations. *Energy and Buildings*, *349*. https://doi.org/10.1016/j.enbuild.2025.116522

Ardigò, L. P., Capelli, C., & Millet, G. P. (2020). Editorial: Human Ultra-Endurance Exercise. In *Frontiers in Physiology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fphys.2020.00664

Belinchón-Demiguel, P., Ruisoto, P., Knechtle, B., Nikolaidis, P. T., Herrera-Tapias, B., & Clemente-Suárez, V. J. (2021). Predictors of athlete's performance in ultra-endurance mountain races. *International Journal of Environmental Research and Public Health*, *18* (3), 1–8. https://doi.org/10.3390/ijerph18030956

Chai, T., & Draxler, R. R. (2014). Root mean square *error* (RMSE) or mean absolute *error* (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7* (3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Costa, R. J. S., Knechtle, B., Tarnopolsky, M., & Hoffman, M. D. (2019). Nutrition for ultramarathon running: Trail, track, and road. *International Journal of Sport Nutrition and Exercise Metabolism*, *29* (2), 130–140. https://doi.org/10.1123/ijsnem.2018-0255

Gregory, D., Kintziger, K., Crouter, S., Sims, C., Kellogg, M., & Fitzhugh, E. (2024). Weather effects on natural surface trail use in an urban wilderness multi-use trail system. *Journal of Outdoor Recreation and Tourism*, *46*, 100757. https://doi.org/10.1016/J.JORT.2024.100757

International *Trail Running* Association. (2022). *ITRA Index Methodology and Race Classification*.

Joyner, M. J. (2017). Physiological limits to endurance exercise performance: influence of sex. In *Journal of Physiology* (Vol. 595, Number 9, pp. 2949–2954). Blackwell Publishing Ltd. https://doi.org/10.1113/JP272268

Knechtle, B., Villiger, E., Valero, D., Braschler, L., Weiss, K., Vancini, R. L., Andrade, M. S., Scheer, V., Nikolaidis, P. T., Cuk, I., Rosemann, T., & Thuany, M. (2024). Analysis of the 10-day ultra-marathon using a predictive XG boost model. *BMC Research Notes*, *17* (1). https://doi.org/10.1186/s13104-024-07028-8

Knechtle, B., Weiss, K., Valero, D., Scheer, V., Villiger, E., Nikolaidis, P. T., Andrade, M., Cuk, I., Gajda, R., Rosemann, T., & Thuany, M. (2025). Change in elevation predicts 100 km ultra marathon performance. *Scientific Reports*, *15* (1). https://doi.org/10.1038/s41598-025-09502-0

Lerebourg, L., Saboul, D., Clémençon, M., & Coquart, J. B. (2022). Prediction of Marathon Performance using Artificial Intelligence. *International Journal of Sports Medicine*, *44* (5), 352–360. https://doi.org/10.1055/a-1993-2371

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2* (1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Molnar, C., Casalicchio, G., & Bischl, B. (2020). *Interpretable Machine learning - - A Brief History, State-of-the-Art and Challenges*. https://doi.org/10.1007/978-3-030-65965-3_28

Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot(2011). Scikit-learn: *Machine learning* in Python. In *Journal of Machine learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Qur'an Kemenag. (2026). Retrieved June 16, 2026, from https://quran.kemenag.go.id/

Shu, D., Wang, J., Zhou, T., Chen, F., Meng, F., Wu, X., Zhao, Z., & Dai, S. (2024). Prediction of half-marathon performance of male recreational marathon runners using nomogram. *BMC Sports Science, Medicine and Rehabilitation*, *16* (1). https://doi.org/10.1186/s13102-024-00889-3

Suter, D., Sousa, C. V., Hill, L., Scheer, V., Nikolaidis, P. T., & Knechtle, B. (2020). Even *Pacing* is associated with faster finishing times in ultramarathon distance *Trail Running*— the "ultra-trail du Mont Blanc" 2008–2019. *International Journal of Environmental Research and Public Health*, *17* (19), 1–11. https://doi.org/10.3390/ijerph17197074

tafsir-ibnu-katsir-surat-al-hasyr. (n.d.).

Turnwald, J., Valero, D., Forte, P., Weiss, K., Villiger, E., Thuany, M., Scheer, V., Wilhelm, M., Andrade, M., Cuk, I., Nikolaidis, P. T., & Knechtle, B. (2025). Analysis of the 50-mile ultramarathon distance using a predictive XGBoost model. *Scientific Reports*, *15* (1). https://doi.org/10.1038/s41598-025-92581-w

UTMB Group. (2023). *UTMB World Series and Race Data Sheet Documentation*.

Wilson, K. J., & Wilson, N. (2024). *Assessing course difficulty and the effect of weather in amateur cross country running races*. http://arxiv.org/abs/2405.09865

LAMPIRAN

Lampiran I. Bukti LOA

Letter of Acceptance

International Journal of Advances in Data and Information Systems

June 11th, 2026

To. 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

We are pleased to inform you that your manuscript, **"Weather-Aware Prediction of Trail Running Finish Times Using Machine Learning"** has been **accepted** for publication in **International Journal of Advances in Data and Information Systems (IJADIS)**, ISSN (Online) 2721-3056 for Vol. 7 No. 2 (2026). All of the accepted manuscripts will go under copyediting as a quality control check before publishing the article.

The IJADIS is indexed in several prestigious databases, including SINTA (Grade 2), Dimensions AI, Google Scholar, Neliti, Portal Garuda, Scilit, BASE, WorldCat, and CrossRef.

Should you have any questions, please do not hesitate to contact the editor of the IJADIS by sending an email to info@ijadis.org.

On behalf of the Editorial Team, we look forward for your future contribution to our journal. Thank you very much in advance for your cooperation.

Yours Sincerely,

Assoc. Prof. Haruna Chiroma, Ph.D.
Editor-in-Chief
International Journal of Advances in Data and Information Systems
Indonesian Scientific Journal

The indexation of the journal

Lampiran II. Jurnal

International Journal of Advances in Data and Information Systems
Vol. 1, No. 2, October 2020, pp. xx~xx
ISSN: 2721-3056, DOI: 10.25008/ijadis.xxxx

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 Jan x, 20xx
Revised Feb x, 20xx
Accepted Apr x, 20xx

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.

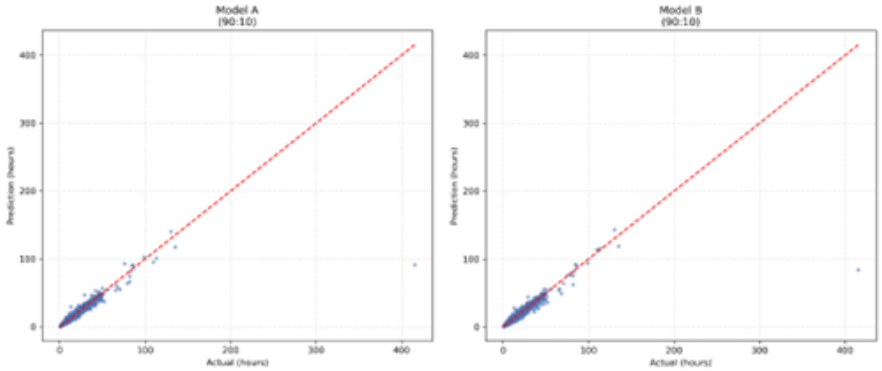


Figure 2. Parity plot demonstrating the relationship between actual and predicted Mean Finish Times (MFT) in hours for Model A and Model B, where the dashed red line represents the ideal $y = x$ reference line.

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

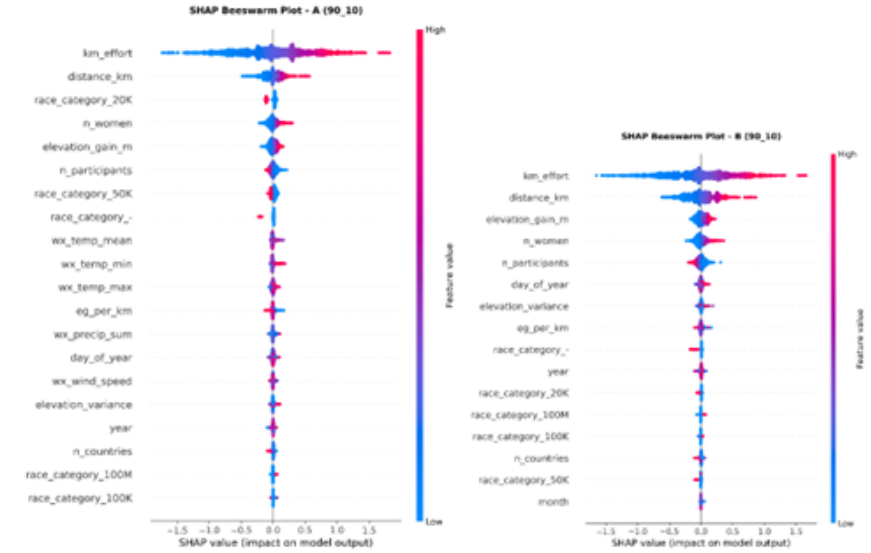


Figure 3. SHAP summary beeswarm plots for Model A and Model B, illustrating the distribution of feature impacts on the model's final prediction output, where red indicates high feature values and blue indicates low feature values.

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- *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.