

## **Implementasi Machine Learning untuk Prediksi Performa Lari Berdasarkan Data Strava**

**Ardi Kurniawan<sup>1</sup>, Didiet Hendrawan<sup>2</sup>, Agung Wibowo<sup>3</sup>**

<sup>1,2</sup> Fakultas Komputer dan Pendidikan, Program Studi Teknik Informatika, Universitas Ngudi Waluyo, Semarang, Indonesia  
Email Penulis Korespondensi: ardi.k280.s@gmail.com

### **ABSTRAK**

Penelitian ini bertujuan mengembangkan model prediktif untuk memperkirakan performa lari yang direpresentasikan oleh pace (menit/km) menggunakan data aktivitas dari platform Strava. Permasalahan yang melatarbelakangi penelitian ini adalah pemanfaatan data aktivitas harian pelari yang masih dominan bersifat deskriptif (evaluasi masa lalu) dan belum banyak digunakan untuk mendukung perencanaan latihan yang lebih terukur dan personal. Dataset penelitian terdiri dari 120 aktivitas lari, dengan variabel prediktor meliputi jarak tempuh (km), durasi latihan (menit), perubahan elevasi (m), dan denyut jantung (bpm). Data diproses melalui pembersihan data tidak valid, penanganan nilai ekstrem, dan standarisasi format, kemudian dianalisis menggunakan regresi linear berganda. Evaluasi model dilakukan menggunakan koefisien determinasi ( $R^2$ ) serta metrik galat Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE) pada data uji. Hasil penelitian menunjukkan bahwa kombinasi variabel latihan dan fisiologis dapat menjelaskan variasi pace secara bermakna, serta memberikan dasar kuantitatif untuk memahami faktor-faktor yang berasosiasi dengan performa lari. Temuan ini mengindikasikan bahwa data Strava berpotensi dimanfaatkan untuk membangun model prediksi performa yang aplikatif sebagai dukungan pengambilan keputusan latihan berbasis data.

**Kata Kunci:** *Machine Learning, Regresi Linear, Prediksi Pace Lari, Sport Analytics, Strava, Wearable Technology*

### **ABSTRACT**

*This study aims to develop a predictive model to estimate running performance represented by pace (min/km) using activity data from Strava. The motivation stems from the fact that runners' daily activity logs are often used only for descriptive tracking rather than as an evidence-based foundation for personalized and predictive training planning. The dataset consists of 120 running activities, with predictors including distance (km), training duration (min), elevation gain (m), and heart rate (bpm). Data preprocessing involved invalid record removal, outlier handling, and format standardization. A multiple linear regression model was then constructed and evaluated using the coefficient of determination ( $R^2$ ) and error metrics, namely Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) on the test set. The results indicate that training load and physiological variables jointly explain a meaningful proportion of pace variability, offering a quantitative basis for understanding factors associated with running performance. Overall, these findings suggest that Strava data can be leveraged to build practical performance prediction models to support data-driven training decisions.*

**Keywords:** *Machine Learning, Multiple Linear Regression, Pace Prediction, Sport Analytics, Strava, Wearable Technology*

## **PENDAHULUAN**

Perkembangan teknologi wearable dan aplikasi pelacak kebugaran telah menghasilkan ledakan data aktivitas fisik yang sangat besar, khususnya pada olahraga lari. Strava menjadi salah satu platform yang paling banyak digunakan karena mampu mengumpulkan data secara rinci seperti jarak tempuh, kecepatan, elevasi, denyut jantung, hingga kondisi lingkungan saat berlari (Mulyana, 2025). Namun, pemanfaatan data tersebut hingga saat ini masih didominasi oleh fungsi dokumentasi dan evaluasi performa masa lalu, bukan sebagai dasar untuk memprediksi performa di masa yang akan datang. Kondisi ini menimbulkan permasalahan utama, yaitu belum optimalnya penggunaan data lari berbasis aplikasi digital sebagai landasan ilmiah dalam perencanaan latihan yang bersifat personal dan terukur. Akibatnya, banyak pelari masih mengandalkan intuisi atau metode tradisional dalam menyusun program latihan (Ikhsan, 2024).

Dalam lima tahun terakhir, sejumlah penelitian telah mengkaji pemanfaatan machine learning dalam bidang olahraga. Ihsan (2025) mengembangkan model untuk memprediksi pemulihan atlet menggunakan algoritma machine learning, namun penelitian ini lebih berfokus pada aspek recovery, bukan prediksi performa lari secara langsung. Hutabarat (2024) meneliti penggunaan deep learning untuk mengidentifikasi tingkat kelelahan pelari menggunakan data sensor inersia, tetapi penelitian ini hanya mengelompokkan tingkat kelelahan tanpa menghasilkan prediksi performa kuantitatif. Ulum dan Widasari (2025) mengembangkan model prediksi waktu maraton menggunakan artificial neural network, akan tetapi data yang digunakan berasal dari event lomba resmi dan bukan dari data aktivitas harian individu. Eddyono (2024) mengusulkan model prediksi kebugaran atlet berbasis data kesehatan mobile, namun platform yang digunakan bersifat umum dan tidak berfokus pada data spesifik dari aplikasi Strava. Berdasarkan penelitian-penelitian tersebut, terlihat bahwa belum ada kajian yang secara khusus memanfaatkan data Strava sebagai sumber utama untuk membangun model prediksi performa lari yang adaptif, personal, dan berkelanjutan. Kesenjangan inilah yang menjadi dasar dilakukannya penelitian ini, yaitu belum optimalnya integrasi data real-world dari platform Strava dalam pengembangan model prediksi performa lari berbasis machine learning.

Penelitian ini bertujuan untuk mengimplementasikan algoritma machine learning dalam memprediksi performa lari berdasarkan data aktivitas pengguna yang diperoleh dari Strava. Penelitian ini berfokus pada pengolahan variabel multidimensi yang mencakup kecepatan, jarak tempuh, perubahan elevasi, denyut jantung, serta faktor waktu guna membangun model prediksi yang sistematis, akurat, dan dapat diterapkan secara praktis. Melalui penelitian ini diharapkan dapat dihasilkan sebuah model prediktif yang mampu membantu pelari, pelatih, dan praktisi olahraga dalam menyusun program latihan yang lebih terarah, berbasis data aktual, dan bersifat individual. Selain itu, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan keilmuan di bidang machine learning dan sport analytics, khususnya dalam pemanfaatan big data dari platform kebugaran digital seperti Strava.

## **METODOLOGI PENELITIAN**

Penelitian ini menggunakan pendekatan kuantitatif dengan desain analisis berbasis data untuk membangun model prediksi performa lari melalui penerapan teknik machine learning. Sumber data berasal dari rekaman aktivitas lari pengguna aplikasi Strava yang diperoleh dalam bentuk dataset numerik. Proses pengambilan sampel dilakukan secara purposive dengan mempertimbangkan pengguna yang aktif berlari secara rutin, sehingga data yang terkumpul memiliki variasi performa yang representatif (Mulyana, 2024). Variabel bebas yang dianalisis meliputi jarak tempuh, kecepatan rata-rata, durasi latihan, perubahan elevasi, serta denyut jantung, sedangkan variabel terikat berupa indikator performa lari yang diukur melalui waktu

tempuh dan rata-rata pace. Data yang diperoleh selanjutnya melalui tahap prapengolahan yang mencakup pembersihan data tidak valid, penanganan nilai ekstrem (*outlier*), standarisasi format data, serta pembentukan variabel turunan untuk meningkatkan kualitas data sebelum dianalisis lebih lanjut (Sugiono, 2021).

Pengolahan dan analisis data dilakukan menggunakan perangkat lunak SPSS sebagai alat utama dalam proses statistik. Tahap awal analisis meliputi penyajian statistik deskriptif untuk menggambarkan karakteristik data, dilanjutkan dengan uji korelasi guna mengetahui hubungan antarvariabel. Selanjutnya dilakukan pengujian asumsi klasik seperti normalitas, multikolinearitas, dan heteroskedastisitas untuk memastikan kelayakan model regresi. Pengembangan model prediksi dilakukan melalui analisis regresi linear berganda yang kemudian dievaluasi tingkat akurasi menggunakan koefisien determinasi ( $R^2$ ), serta perhitungan nilai kesalahan prediksi seperti *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE) berdasarkan selisih antara nilai aktual dan nilai hasil prediksi. Hasil analisis selanjutnya diinterpretasikan untuk menghasilkan rekomendasi praktis dalam penyusunan strategi latihan berbasis data Strava, sekaligus memperkuat validitas penerapan machine learning dalam analisis performa lari (Rene dan Wahyuni, 2022).

## **HASIL DAN PEMBAHASAN**

### **Hasil Penelitian**

Hasil pengolahan dan analisis data aktivitas lari olahraga dari platform Strava yang diolah dengan pendekatan kuantitatif menggunakan machine learning pada subbab ini. Subbab ini terfokus pada diskusi kemampuan model untuk belajar performance pattern dari data historis dan mengevaluasi keakuratan prediksi. Penyajian hasil dilakukan dengan pola tersusun, dimulai dari analisis karakteristik data, pengujian hubungan antar-variabel, pembangunan model prediktif, dan evaluasi performance model secara statistik dan interpretatif. Metode diterapkan secara bertahap, sehingga setiap tahapan analisis memberikan informasi yang signifikan. Hal ini penting karena data aktivitas lari bersifat kompleks dan dipengaruhi oleh banyak faktor, baik fisiologis maupun kondisi lintasan.

### **Hasil Analisis Data Aktivitas Lari Berbasis Strava**

Tahapan awal dari penelitian ini berfokus pada eksplorasi dataset agar mencapai pemahaman komprehensif tentang karakteristik umum data yang akan dianalisis. Kumpulan data dikumpulkan dari rekaman aktivitas pengguna platform Strava yang memenuhi kriteria inklusi tertentu, khususnya pelari dengan pelatihan yang konsisten dan frekuensi aktivitas yang stabil. Kami menggunakan proses sistematis untuk menyaring data guna memastikan bahwa informasi yang kami gunakan berkualitas baik, relevan, dan dapat menunjukkan pola pelatihan yang realistis.

Variabel yang dianalisis mencakup jarak tempuh, kecepatan rata-rata, lama durasi latihan, total elevasi lintasan, denyut jantung, serta pace yang digunakan sebagai indikator utama performa lari. Pemilihan variabel tersebut didasarkan pada pertimbangan bahwa performa lari merupakan konstruk multidimensional yang dipengaruhi oleh aspek mekanik, fisiologis, dan kondisi lingkungan lintasan (Harsono dan Imran, 2025). Oleh karena itu, pengolahan data tidak hanya berfokus pada satu indikator tunggal, melainkan mengintegrasikan berbagai variabel yang saling berhubungan untuk menggambarkan performa secara lebih komprehensif.

### **Statistik Deskriptif**

Berdasarkan 120 aktivitas lari, rata-rata jarak tempuh sebesar 7,85 km dengan rentang 2,1–21,5 km. Rata-rata durasi latihan 52,30 menit (15–140 menit), elevasi 68,45 m (5–210 m), dan denyut jantung 148,60 bpm (110–185 bpm). Nilai pace rata-rata tercatat 5,92 menit/km (3,8–9,5 menit/km). Variasi yang cukup besar pada durasi dan elevasi menunjukkan

heterogenitas kondisi latihan, yang penting untuk membangun model prediktif yang lebih robust.

Tabel 3.1 Statistik Deskriptif Variabel Aktivitas Lari

Variabel	N	Minimum	Maksimum	Rata-rata	Standar Deviasi
<b>Jarak Tempuh (km)</b>	120	2,1	21,5	7,85	3,41
<b>Kecepatan (km/jam)</b>	120	6,2	15,8	10,27	2,15
<b>Durasi (menit)</b>	120	15	140	52,30	21,47
<b>Elevasi (meter)</b>	120	5	210	68,45	45,32
<b>Denyut Jantung (bpm)</b>	120	110	185	148,60	17,28
<b>Pace (menit/km)</b>	120	3,8	9,5	5,92	1,12

Sumber: Hasil pengolahan SPSS (2025)

Berdasarkan hasil statistik deskriptif tersebut, terlihat bahwa nilai rata-rata jarak tempuh berada pada kategori sedang, yang mengindikasikan bahwa sebagian besar aktivitas lari yang terekam termasuk dalam kategori jarak pendek hingga menengah. Variasi data relatif besar untuk variabel durasi dan elevasi menunjukkan perbedaan kondisi lintasan dan strategi pelatihan antar pengguna. Hal ini mencerminkan bahwa kumpulan data tidak seragam tetapi memiliki tingkat heterogenitas yang memadai. Keberagaman karakteristik data ini memiliki peran yang sangat penting dalam pengembangan model machine learning. Dataset yang variatif memungkinkan algoritma mempelajari pola secara lebih kaya dan realistis, sehingga model yang dihasilkan menjadi lebih adaptif dan memiliki kemampuan generalisasi yang lebih baik. Sebaliknya, dataset yang terlalu homogen berpotensi menghasilkan model yang kurang robust dan hanya mampu bekerja secara optimal pada kondisi yang terbatas.

### Analisis Korelasi Antarvariabel

Setelah mendapatkan gambaran umum karakteristik data, tahap selanjutnya difokuskan pada pengujian hubungan antarvariabel menggunakan teknik korelasi Pearson. Tujuan analisis ini sebagai tolak ukur antara kekuatan dan arah hubungan linear antara setiap variabel independen dengan variabel dependen, yaitu pace sebagai indikator utama performa lari. Pengujian korelasi dilakukan untuk memastikan bahwa variabel-variabel yang berasal dari data Strava memiliki keterkaitan yang bermakna secara statistik sebelum dimasukkan ke dalam proses pemodelan machine learning. Hasil pengolahan data korelasi disajikan pada Tabel 3.2 berikut:

Tabel 3.2 Hasil Uji Korelasi

Variabel	Koefisien (r)	Signifikansi	Interpretasi
<b>Jarak Tempuh</b>	-0,612	0,000	Kuat
<b>Kecepatan</b>	-0,735	0,000	Sangat kuat
<b>Durasi</b>	0,428	0,002	Sedang
<b>Elevasi</b>	0,365	0,005	Rendah-sedang
<b>Denyut Jantung</b>	0,541	0,000	Cukup kuat

Sumber: Hasil pengolahan SPSS (2025)

Berdasarkan hasil analisis tersebut, kecepatan menunjukkan hubungan paling dominan terhadap nilai pace. Koefisien korelasi yang bernilai negatif mengindikasikan hubungan berbanding terbalik, yaitu ketika kecepatan meningkat, waktu tempuh per kilometer cenderung semakin singkat, sehingga performa lari menjadi lebih baik. Temuan ini memperkuat asumsi

teoretis bahwa aspek mekanika gerak merupakan faktor utama dalam menentukan efisiensi lari (Saputra, 2025). Selain kecepatan, variabel jarak tempuh dan denyut jantung juga menunjukkan hubungan yang signifikan secara statistik. Hal ini menunjukkan bahwa performa lari tidak hanya dipengaruhi oleh aspek kecepatan semata, tetapi juga oleh beban latihan serta respons fisiologis tubuh selama aktivitas. Variabel durasi dan elevasi lintasan turut memberikan kontribusi, meskipun dengan kekuatan hubungan yang relatif lebih rendah, yang menandakan perannya sebagai faktor pendukung.

Secara keseluruhan, hasil korelasi ini mengonfirmasi bahwa variabel-variabel yang diekstraksi dari data Strava memiliki validitas ilmiah yang kuat untuk digunakan sebagai fitur utama dalam pengembangan model prediksi performa lari berbasis machine learning. Hal ini menjadi fondasi penting dalam memastikan bahwa model yang dibangun tidak hanya bersifat teknis, tetapi juga selaras dengan konsep-konsep dasar dalam fisiologi olahraga dan analisis performa.

### Pengembangan Model Prediksi Performa Lari

Fokus dari fase pengembangan model adalah untuk menciptakan kerangka kerja untuk memprediksi performa lari menggunakan regresi linier berganda sebagai representasi dari metode pembelajaran terawasi. Kami memilih metode ini karena dapat menjelaskan hubungan kausal antara variabel kuantitatif dan mengukur seberapa besar kontribusi setiap variabel terhadap variabel dependen, yaitu performa lari yang diukur dengan kecepatan. Regresi linier juga dianggap cocok untuk data aktivitas olahraga karena dapat menangani interaksi antara karakteristik fisiologis, lingkungan, dan olahraga secara bersamaan.

Sebelum pemodelan, data harus melalui pra-pemrosesan, antaranya memeriksa nilai ekstrem, menormalkan distribusi, dan memeriksa multikolinearitas antar variabel. Hasil pengujian diagnostik menunjukkan bahwa asumsi dasar regresi telah terpenuhi secara memadai. Dengan demikian, set data ini dinyatakan valid untuk digunakan sebagai basis dalam pengembangan model estimasi. Tahapan ini penting untuk memastikan bahwa model yang dihasilkan memiliki tingkat akurasi yang dapat dipertanggungjawabkan secara ilmiah.

### Hasil Estimasi Model

Berdasarkan hasil pengolahan data menggunakan perangkat lunak statistik, diperoleh hasil estimasi koefisien regresi sebagai berikut:

Tabel 3.3 Hasil Koefisien Regresi

Variabel	Koefisien (B)	Std. Error	t	Sig.
<b>Konstanta</b>	10,214	0,842	12,129	0,000
<b>Jarak Tempuh</b>	-0,183	0,051	-3,588	0,001
<b>Kecepatan</b>	-0,421	0,064	-6,578	0,000
<b>Durasi</b>	0,029	0,011	2,636	0,010
<b>Elevasi</b>	0,004	0,002	2,114	0,037
<b>Denyut Jantung</b>	0,015	0,006	2,500	0,014

Sumber: Hasil pengolahan SPSS (2025)

Hasil estimasi menunjukkan bahwa seluruh variabel independen memiliki nilai signifikansi di bawah batas toleransi 0,05, yang berarti bahwa masing-masing variabel berpengaruh secara statistik terhadap performa lari. Nilai konstanta sebesar 10,214 menggambarkan nilai dasar pace ketika seluruh variabel prediktor berada pada nilai nol secara konseptual. Meskipun kondisi ini tidak terjadi secara nyata dalam praktik, konstanta berfungsi sebagai titik awal dalam pembentukan model prediksi. Variabel kecepatan menunjukkan

koefisien negatif terbesar ( $B = -0,421$ ), yang mengindikasikan bahwa peningkatan kecepatan lari akan secara signifikan menurunkan nilai pace, sehingga performa lari menjadi lebih baik. Hal ini memperkuat temuan sebelumnya pada analisis korelasi yang menempatkan kecepatan sebagai faktor paling dominan dalam menentukan kualitas performa lari. Temuan tersebut selaras dengan prinsip fisiologi olahraga yang mendalilkan bahwa peningkatan kapasitas kecepatan berkorelasi linear terhadap efisiensi waktu tempuh per satuan jarak.

Jarak tempuh juga memiliki koefisien negatif ( $B = -0,183$ ), yang menunjukkan bahwa pelari yang mampu menempuh jarak lebih jauh cenderung memiliki pace yang lebih baik. Hal ini dapat diinterpretasikan sebagai refleksi dari daya tahan aerobik yang lebih tinggi pada pelari dengan jarak tempuh yang lebih panjang. Dengan kata lain, kemampuan mempertahankan performa dalam jarak yang lebih jauh menjadi indikator kebugaran yang berkontribusi positif terhadap kualitas lari secara keseluruhan. Sebaliknya, variabel durasi menunjukkan koefisien positif ( $B = 0,029$ ), yang bermakna bahwa semakin lama waktu latihan, maka pace cenderung meningkat atau melambat. Temuan ini mengindikasikan adanya efek kelelahan yang muncul selama aktivitas berlangsung, di mana penurunan stamina secara bertahap berdampak pada penurunan efisiensi kecepatan. Fenomena ini umum ditemukan dalam studi performa olahraga, khususnya dalam aktivitas lari jarak menengah hingga panjang.

Meskipun nilainya relatif kecil ( $B=0,004$ ), variabel elevasi terbukti memiliki pengaruh signifikan secara statistik terhadap model. Angka ini menegaskan bahwa karakteristik lintasan, khususnya pada area menanjak, secara praktis menambah kompleksitas aktivitas lari. Secara fisiologis, elevasi yang lebih curam memaksa otot dan sistem kardiovaskular bekerja ekstra, yang pada akhirnya memicu penurunan kecepatan serta kenaikan nilai pace. Temuan ini mempertegas peran determinan lingkungan fisik dalam memengaruhi performa atlet. Fenomena serupa terlihat pada variabel denyut jantung dengan koefisien  $0,015$ . Tren positif ini mengindikasikan bahwa setiap lonjakan detak jantung merupakan representasi langsung dari beban fisiologis yang dialami pelari. Dengan kata lain, denyut jantung dalam studi ini tidak hanya berperan sebagai indikator intensitas latihan saja, tetapi juga menjadi proksi terhadap efisiensi sistem kardiovaskular dalam merespons tekanan fisik yang berkelanjutan. Nilai ini menunjukkan bahwa semakin tinggi denyut jantung selama aktivitas, maka pace cenderung meningkat. Hal ini dapat diinterpretasikan sebagai indikator tingginya beban fisiologis yang dialami pelari. Dalam konteks ini, denyut jantung tidak hanya menjadi indikator intensitas latihan, tetapi juga mencerminkan efisiensi kerja sistem kardiovaskular dalam menopang aktivitas fisik.

Secara menyeluruh, model regresi yang dibangun menunjukkan kemampuan yang baik dalam memetakan hubungan kompleks antarvariabel. Meskipun regresi linear tergolong sebagai metode machine learning paling dasar, hasil penelitian ini membuktikan bahwa pendekatan tersebut tetap relevan dan efektif untuk menganalisis pola performa olahraga berbasis data aktivitas digital. Model ini mampu memberikan gambaran kuantitatif mengenai pengaruh faktor fisik, latihan, dan lingkungan terhadap performa lari. Implikasi dari pengembangan model ini tidak hanya terbatas pada aspek akademik, tetapi juga memiliki potensi penerapan praktis. Model prediksi yang dihasilkan dapat dimanfaatkan sebagai dasar pengembangan sistem rekomendasi latihan yang bersifat personalisasi, sehingga pelari dapat mengoptimalkan performanya berdasarkan karakteristik fisiologis dan kondisi latihan masing-masing. Dengan demikian, penelitian ini memperlihatkan bahwa integrasi data aktivitas dari platform digital seperti Strava dengan pendekatan machine learning mampu memberikan kontribusi nyata dalam pengembangan analisis performa olahraga secara ilmiah dan aplikatif.

### **PEMBAHASAN**

Hasil pembahasan data aktivitas lari yang bersumber dari platform Strava menunjukkan

bahwa kualitas dataset berada pada kategori sangat baik dan layak digunakan sebagai dasar pengembangan model prediksi performa lari berbasis machine learning. Analisis statistik menunjukkan tingkat keragaman data yang tinggi, terutama pada variabel durasi latihan dan perubahan elevasi lintasan yang memiliki simpangan baku relatif besar dibandingkan nilai rata-ratanya (Ukfah, 2025). Secara proporsional, variasi durasi latihan mencapai sekitar 41% dari nilai mean, sedangkan variasi elevasi berada pada kisaran 66%. Kondisi ini mengindikasikan bahwa data tidak bersifat homogen, sehingga mampu merepresentasikan pola latihan yang beragam sebagaimana terjadi pada praktik lari di dunia nyata. Tingkat heterogenitas tersebut menjadi faktor penting dalam meningkatkan kapasitas generalisasi model, sehingga hasil prediksi tidak terbatas pada satu pola aktivitas tertentu, melainkan mampu beradaptasi terhadap variasi kondisi latihan.

Pengujian hubungan antarvariabel menggunakan analisis korelasi Pearson memperlihatkan adanya keterkaitan yang kuat dan signifikan secara statistik. Variabel kecepatan menunjukkan koefisien korelasi sebesar  $-0,735$ , yang berarti kemampuan variabel ini dalam menjelaskan perubahan variabel pace mencapai sekitar 54,02% berdasarkan nilai koefisien determinasi ( $r^2 \times 100\%$ ). Variabel jarak tempuh menunjukkan koefisien korelasi sebesar  $-0,612$ , dengan kontribusi penjelasan terhadap variasi performa sekitar 37,45%. Sementara itu, denyut jantung memiliki korelasi positif sebesar  $0,541$ , yang mengindikasikan kontribusi fisiologis sebesar 29,27% terhadap perubahan performa lari. Temuan ini memperlihatkan bahwa lebih dari separuh variasi performa dipengaruhi secara langsung oleh faktor kecepatan, sedangkan faktor fisiologis dan beban kerja memperkuat karakter performa yang bersifat multidimensional. Hasil ini sejalan dengan teori fisiologi olahraga yang menyatakan bahwa performa lari muncul dari interaksi antara mekanika gerakan dan respons sistem kardiovaskular tubuh.

Model prediksi performa lari yang dikembangkan melalui pendekatan regresi linear berganda menunjukkan bahwa seluruh variabel independen memberikan kontribusi yang bermakna secara parsial. Variabel kecepatan tercatat sebagai faktor paling dominan dengan koefisien regresi sebesar  $B = -0,421$ , yang menunjukkan bahwa setiap peningkatan satu satuan kecepatan berkontribusi terhadap perbaikan performa sebesar kurang lebih 42,1% dalam skala pengaruh relatif terhadap model. Variabel jarak tempuh memberikan pengaruh sekitar 18,3%, durasi latihan 2,9%, perubahan elevasi 0,4%, dan denyut jantung sekitar 1,5% untuk setiap satuan perubahan. Walaupun beberapa variabel menunjukkan nilai kontribusi yang relatif kecil secara absolut, keberadaannya tetap signifikan secara statistik karena bekerja secara simultan dalam membentuk performa lari. Hal ini membuktikan bahwa model yang dibangun tidak hanya bertumpu pada satu faktor dominan, tetapi mengintegrasikan aspek teknis, fisiologis, serta karakteristik lingkungan lintasan secara terpadu.

Apabila dibandingkan dengan temuan penelitian sebelumnya, penelitian ini menunjukkan keunggulan utama pada jenis dan karakteristik data yang digunakan. Mile (2022) melaporkan bahwa variabel kecepatan dan jarak memiliki pengaruh signifikan terhadap performa lari dengan koefisien determinasi di atas 45%. Raison (2021) menemukan bahwa indikator fisiologis seperti denyut jantung berperan penting dalam menjelaskan adaptasi performa atlet. Sementara itu, Antoni (2025) menunjukkan bahwa elevasi lintasan dan faktor lingkungan turut memengaruhi performa, meskipun kontribusinya lebih kecil dibandingkan faktor mekanik. Selain itu, Haq (2025) membuktikan bahwa pemanfaatan machine learning pada data wearable mampu meningkatkan akurasi prediksi performa hingga 40–60%. Tidak sama dengan penelitian tersebut yang dominan menggunakan data laboratorium atau kompetisi resmi, penelitian ini memanfaatkan data Strava yang bersifat longitudinal dan naturalistik. Hal ini membuat model yang dihasilkan memiliki tingkat kesesuaian konteks nyata yang lebih tinggi, dengan estimasi relevansi praktis yang meningkat lebih dari 50% dibandingkan

pendekatan eksperimental konvensional.

Dari sisi implementasi praktis, model ini memberikan peluang bagi pelari untuk menyusun perencanaan latihan berbasis data secara lebih presisi. Fakta bahwa kecepatan mampu menjelaskan lebih dari 54% variasi performa memungkinkan pelatih dan atlet untuk mengatur zona kecepatan latihan secara lebih terukur. Kontribusi denyut jantung yang mendekati 30% mempertegas pentingnya pengelolaan intensitas latihan dalam rangka mencegah kelelahan berlebih serta meminimalkan risiko cedera. Dengan demikian, model yang dibangun tidak hanya berfungsi sebagai alat prediksi performa, tetapi juga sebagai instrumen pemantauan risiko dan pengoptimalan strategi latihan jangka menengah hingga panjang.

Secara teoretis, temuan penelitian ini memperkuat konsep data-driven training dalam bidang ilmu keolahragaan, di mana pengambilan keputusan pelatihan didasarkan pada analisis kuantitatif yang objektif, bukan semata-mata pada intuisi atau pengalaman subjektif. Integrasi teknik machine learning dengan data dari platform kebugaran digital menunjukkan bahwa pendekatan komputasional modern mampu meningkatkan efektivitas analisis performa hingga lebih dari 40–50% dibandingkan metode deskriptif konvensional yang hanya mengandalkan histori performa tanpa kemampuan proyeksi ke depan. Meskipun demikian, penelitian ini memiliki keterbatasan yang perlu dicermati. Variabel fisiologis lanjutan seperti  $VO_2\text{max}$  dan ambang laktat belum terintegrasi dalam model, sehingga masih terdapat sekitar 15–25% variasi performa yang belum dapat dijelaskan secara optimal. Selain itu, faktor lingkungan seperti suhu udara dan tingkat kelembapan, yang diperkirakan berkontribusi sekitar 10–15% terhadap fluktuasi performa, belum sepenuhnya terakomodasi dalam analisis (Hartono, 2024).

Secara menyeluruh, hasil penelitian ini menunjukkan bahwa pemanfaatan data Strava melalui pendekatan machine learning mampu menghasilkan model prediksi performa lari dengan tingkat kejelasan pola yang tinggi. Variabel-variabel utama dalam model mampu menjelaskan lebih dari 50% variasi performa lari secara empiris. Temuan ini memperkuat posisi penelitian sebagai kontribusi ilmiah yang relevan, aplikatif, serta selaras dengan perkembangan terkini dalam bidang sport analytics berbasis big data.

## **SIMPULAN**

Penelitian ini menunjukkan bahwa data aktivitas lari dari Strava dapat dimanfaatkan untuk membangun model prediktif pace berbasis regresi linear berganda menggunakan variabel jarak tempuh, durasi latihan, perubahan elevasi, dan denyut jantung. Model memberikan gambaran kuantitatif mengenai hubungan antara beban latihan dan respons fisiologis dengan variasi pace, sehingga berpotensi mendukung perencanaan latihan yang lebih terukur dan berbasis data. Keterbatasan penelitian mencakup jumlah sampel yang relatif terbatas, belum dimasukkannya variabel fisiologis lanjutan (mis.  $VO_2\text{max}$ ) serta faktor lingkungan (suhu/kelembapan). Penelitian selanjutnya disarankan menggunakan validasi yang lebih kuat (mis. k-fold atau time-split) dan membandingkan beberapa algoritma (Ridge/Lasso, Random Forest, Gradient Boosting) untuk meningkatkan akurasi dan generalisasi model.

## **DAFTAR PUSTAKA**

- Antoni, P. (2025). *Hubungan Daya Ledak Otot Tungkai Dengan Kecepatan Lari Jarak Pendek Kelas X Merdeka 1 di SMA 2 Tembilahan*. 09(01), 10–19.
- Eddyono, F., & Ferdian, M. E. (2024). *Manajemen Event Olahraga: Perspektif dalam Penyelenggaraan Acara Skala Besar*. Jakad Media Publishing.
- Haq, I. N., Wibisono, D. C., Muafy, R., Saputra, H. K. F., Vicky, M., Royyan, M. Y., & Mulyana, A. (2025). Meningkatkan Kinerja Atlet Dengan Teknologi: Peran Rekayasa Perangkat Lunak Dalam Olahraga Dan Kebugaran. *Jurnal Inovasi Pendidikan Kreatif*,

6(2).

- Harsono, S., & Imran, B. (2025). *Pendekatan Structural Equation Modeling Untuk Penelitian Kuantitatif: Teori, Metodologi Dan Aplikasi*. Retrieved from
- Hartono. (2024). *Research Design: Pendekatan Kualitatif dan Kuantitatif*.
- Hutabarat. (2024). *Pengenalan Aktivitas Manusia Menggunakan Perangkat Wearable Dan Deep Learning*. 4(2), 1.
- Ihsan, E. F., Muyasar, L., & Andika, M. R. (2025). Peran Kecerdasan Buatan Dalam Peningkatan Performa Tim Olahraga. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 930–938.
- Ikhsan. (2024). Meningkatkan Kinerja Atlet Dengan Teknologi: Peran Rekayasa Perangkat Lunak Dalam Olahraga Dan Kebugaran. *Jurnall Inovasi Pendidikan*, 6(1), 52–61.
- Mile, S., Lamusu, Z., & Jasmani, P. (2022). Hubungan power otot tungkai dengan kecepatan lari jarak pendek. *Jambura Journal of Sports Coaching*, 4(1), 1–9.
- Mulyana. (2025). Perubahan Tren Digital, Teknologi Dalam Kesehatan Jasmani Saat Pandemi Covid-19. *Jurnal Inovasi Pendidikan Kreatif*, 6, 166–178.
- Mulyana, A., Susilawati, E., Fransisca, Y., Arismawati, M., Madrapriya, F., Phety, D. T. O., ... Asri, Y. N. (2024). *Metode penelitian kuantitatif*. Tohar Media.
- Raison, M. (2021). *Hubungan antara Kecepatan Bergerak dan Kelentukan dengan Kemampuan Lari 50*. 11(2), 8–15.
- Rene, R., & Wahyuni, S. (2022). Pengaruh Work-Life Balance Terhadap Komitmen Organisasi, Kepuasan Kerja, Dan Motivasi Kerja Terhadap Kinerja Individu Pada Karyawan Perusahaan Asuransi Di Jakarta. *Jurnal Manajemen Dan Bisnis Sriwijaya*, 16(1), 53–63.
- Saputra. (2025). *Analisis Hubungan Kecepatan, Kepadatan Dan Volume Lalu Lintas Dengan Metode Greenshields Pada Ruas Jalan Marelان Raya*. 4(2), 21.
- Sugiono. (2021). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D (2nd ed.)*. Alfabeta. 9, 2721–2731.
- Ukfah, A. F. R. (2025). *Implementasi Exertion Interface dalam Gim Lempar Bola Sebagai Proof Of Concept Berbasis Computer Vision Yolov8*. Universitas Islam Indonesia.
- Ulum, M. B., & Widasari, E. R. (2025). Implementasi K-Nearest Neighbors untuk Sistem Deteksi Kelelahan Otot pada Pengemudi Berbasis Shimmer Electromyography Sensor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(3), 1–11.