

## Analisis Faktor yang Mempengaruhi *Pace* Lari Mahasiswa Menggunakan *Random Forest*

Dihin Muriyatmoko<sup>1</sup>, Miftahuddin Fahmi<sup>2</sup>, Royhan Akbar<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor

E-mail: \*<sup>1</sup>[dihin@unida.gontor.ac.id](mailto:dihin@unida.gontor.ac.id), <sup>2</sup>[miftahuddinfahmi@unida.gontor.ac.id](mailto:miftahuddinfahmi@unida.gontor.ac.id),

<sup>3</sup>[royhanakbar38@student.cs.unida.gontor.ac.id](mailto:royhanakbar38@student.cs.unida.gontor.ac.id)

**Abstract** – Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi *pace* lari mahasiswa Universitas Darussalam menggunakan algoritma *Random Forest*. Variabel yang dianalisis meliputi semester, usia, tinggi badan, berat badan, indeks massa tubuh, frekuensi latihan per minggu, dan status perokok. Data penelitian dikumpulkan melalui kuesioner dengan total 118 data responden. Untuk meningkatkan jumlah dan variasi data, dilakukan proses augmentasi dengan menambahkan noise ringan pada variabel *pace*, sehingga diperoleh 690 data. Sebelum digunakan dalam proses pelatihan model, data melalui tahap prapemrosesan yang meliputi konversi tipe data ke format numerik serta pembersihan data dari nilai hilang. Hasil pemodelan regresi menunjukkan nilai koefisien determinasi ( $R^2$ ) sebesar 0,995 dan nilai root mean squared error (RMSE) sebesar 0,255. Analisis *feature importance* pada algoritma *Random Forest* menunjukkan bahwa indeks massa tubuh merupakan faktor yang paling berpengaruh terhadap *pace* lari mahasiswa. Hasil penelitian ini diharapkan dapat menjadi bahan evaluasi dan tolak ukur bagi mahasiswa dalam meningkatkan performa *pace* lari, serta dapat dijadikan referensi bagi penelitian selanjutnya di bidang analisis performa olahraga berbasis data.

**Kata Kunci** — *Feature Importance, Mahasiswa, Pace Lari, Random Forest*

### 1. PENDAHULUAN

Lari merupakan salah satu aktivitas gerak dasar manusia yang ditandai dengan pergerakan kaki secara bergantian ke depan dengan kecepatan yang lebih tinggi dibandingkan berjalan. Dalam aktivitas berlari, terdapat fase melayang, yaitu kondisi ketika seluruh tubuh terangkat dari permukaan tanah sehingga tidak ada bagian kaki yang menyentuh tanah [1].

Anak muda, khususnya mahasiswa Universitas Darussalam Gontor, sering kali kurang menyadari pentingnya mengukur dan mengevaluasi *pace* lari secara objektif [2][3]. Akibatnya, meskipun aktivitas lari dilakukan secara rutin, perkembangan performa lari yang dicapai belum optimal dan sulit diukur secara akurat.

Dalam upaya untuk meningkatkan kualitas dan performa mahasiswa dalam bidang olahraga, khususnya cabang lari, diperlukan adanya pemahaman yang mendalam mengenai berbagai aspek yang memengaruhi kemampuan fisik mahasiswa. Namun, dalam kenyataannya masih terdapat beberapa permasalahan yang dihadapi, antara lain Tidak adanya data yang dapat dijadikan sebagai tolak ukur bagi mahasiswa Universitas Darussalam untuk meningkatkan *pace*.

Di sisi lain, perkembangan teknologi serta ketersediaan data dalam bidang olahraga telah membuka peluang yang signifikan untuk menganalisis faktor-faktor yang memengaruhi *pace* lari melalui pendekatan analisis berbasis data. Dengan analisis yang tepat, mahasiswa dapat memahami variabel-variabel yang memiliki pengaruh dominan terhadap kecepatan lari mereka, sehingga dapat dijadikan dasar dalam perancangan program latihan yang efektif, terukur, dan bersifat personal.

Berbagai penelitian sebelumnya menunjukkan bahwa performa lari, khususnya kecepatan lari jarak pendek, dipengaruhi oleh beragam faktor fisik dan biomekanik. Parwata [4] menemukan adanya hubungan yang signifikan antara tinggi badan dan berat badan terhadap kecepatan lari 100 meter pada mahasiswa. Penelitian lain oleh Padaningpuri dan Andar [5] menunjukkan bahwa faktor-faktor kinesiology memiliki kontribusi yang nyata terhadap kecepatan lari mahasiswa. Selain itu, Quoc [6] mengungkapkan bahwa perkembangan kecepatan lari mahasiswa non-spesialis dipengaruhi oleh kombinasi faktor antropometri, kondisi fisik, dan pola latihan. Hasan dan Sukanto [7] juga melaporkan bahwa kecepatan reaksi kaki dan kelentukan memberikan kontribusi yang signifikan terhadap kemampuan lari 100 meter pada mahasiswa.

Selain faktor-faktor tersebut, Ikadarny dan Karim [8] mengungkapkan bahwa kecepatan lari dan daya ledak tungkai memiliki peran penting dalam menunjang kemampuan gerak eksplosif, yang secara tidak langsung berkaitan dengan performa lari. Temuan ini memperkuat bahwa komponen kecepatan dan kekuatan otot tungkai merupakan aspek fundamental dalam peningkatan performa aktivitas lari.

Meskipun berbagai penelitian telah membahas faktor-faktor yang memengaruhi performa lari, masih terdapat keterbatasan dalam penyediaan data yang kontekstual dan spesifik bagi mahasiswa Universitas Darussalam Gontor. Oleh karena itu, diperlukan kajian yang lebih terarah untuk menganalisis faktor-faktor yang memengaruhi pace lari mahasiswa di lingkungan Universitas Darussalam Gontor.

Penelitian ini diharapkan dapat memberikan kontribusi ilmiah sebagai referensi bagi penelitian selanjutnya serta memberikan manfaat praktis bagi mahasiswa dalam meningkatkan performa lari mereka secara objektif dan berbasis data.

## 2. METODE PENELITIAN

Metode yang diterapkan dalam penelitian ini adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Sejak diperkenalkan pada tahun 1996, CRISP-DM dikembangkan sebagai kerangka kerja standar yang banyak digunakan untuk mendukung proses penambangan data. Metode ini terdiri atas enam tahapan utama, yaitu business understanding, data understanding, data preparation, modelling, evaluation, dan deployment.



Gambar 1. *Crisp-Dm*

### 2.1 *Businesses Understanding*

Tahap business understanding bertujuan untuk mengidentifikasi dan memahami tujuan penelitian secara komprehensif [9]. Permasalahan yang dihadapi adalah belum optimalnya pemahaman mengenai faktor-faktor yang memengaruhi *pace* lari mahasiswa Universitas Darussalam secara objektif dan terukur. Meskipun aktivitas lari telah menjadi bagian dari gaya hidup dan kegiatan olahraga mahasiswa, evaluasi performa lari masih sering dilakukan secara subjektif tanpa dukungan analisis data yang komprehensif. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan mengidentifikasi faktor-faktor yang berpengaruh terhadap *pace* lari mahasiswa menggunakan teknik data mining, sehingga hasil yang diperoleh dapat menjadi dasar dalam perencanaan latihan yang lebih efektif serta peningkatan performa lari mahasiswa.

### 2.2 *Data Understanding*

Tahap data understanding merupakan fase awal dalam proses data mining yang berfokus pada kegiatan pengumpulan, eksplorasi, serta evaluasi data guna memperoleh pemahaman menyeluruh terhadap karakteristik, struktur, dan kualitas data yang tersedia. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan memiliki relevansi, tingkat keandalan yang memadai, serta kelayakan untuk dianalisis pada tahap selanjutnya [10]. Data dalam penelitian ini diperoleh dari formulir pengisian yang dilakukan oleh mahasiswa pada periode 2025–2026, kemudian disimpan dalam format Microsoft Excel. Adapun variabel-variabel yang terkandung dalam data tersebut meliputi:

- a. Semester
- b. Usia
- c. Berat badan
- d. Tinggi badan
- e. Index massa tubuh

- f. Latihan perminggu

### 2.3 Data preparation

Tahap *data preparation* mencakup serangkaian proses yang bertujuan untuk menghasilkan *dataset* akhir yang siap digunakan pada tahap *modeling* [11]. Pada tahap ini, langkah awal yang dilakukan adalah pembersihan data (*data cleaning*), yang meliputi proses pemeriksaan data untuk mengidentifikasi serta menangani kesalahan pencatatan, duplikasi data, dan keberadaan nilai yang hilang. Selanjutnya, dilakukan proses transformasi data, yaitu penyesuaian format data, pengkodean variabel, serta penambahan atau pembentukan atribut baru sesuai dengan kebutuhan dan tujuan analisis, sehingga data yang dihasilkan dapat mendukung proses pemodelan secara optimal.

### 2.4 Modeling

Pada tahap modeling, proses data mining dilakukan dengan menerapkan algoritma yang telah ditentukan guna menghasilkan model yang sesuai dengan tujuan analisis [12]. Dalam penelitian ini, algoritma yang digunakan adalah *Random Forest Regressor*, yang berfungsi untuk memodelkan hubungan antara sejumlah variabel independen dengan variabel dependen berupa *pace* lari mahasiswa. *Random Forest Regressor* merupakan metode ensemble learning yang bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksinya untuk memperoleh estimasi yang lebih akurat dan stabil.

Selain itu, *Random Forest Regressor* juga memiliki kemampuan yang baik dalam menangani data berdimensi tinggi serta mengurangi risiko overfitting melalui mekanisme pengacakan data dan fitur pada setiap pohon yang dibangun [13]. Dengan demikian, penggunaan algoritma ini diharapkan mampu menghasilkan model prediksi *pace* lari yang optimal dan sesuai dengan kebutuhan analisis penelitian.

### 2.5 Evaluasi

Pada tahap *evaluation*, model yang telah dibangun dievaluasi untuk menilai kualitas serta kinerjanya dalam mencapai tujuan penelitian. Tahap ini bertujuan untuk memvalidasi sejauh mana model *Random Forest Regressor* mampu memberikan prediksi *pace* lari mahasiswa secara akurat dan andal.

Evaluasi model dalam penelitian ini dilakukan dengan menggunakan beberapa metrik kinerja regresi, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), dan koefisien determinasi ( $R^2$ ). MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, sedangkan MSE memberikan penalti yang lebih besar terhadap kesalahan prediksi yang bernilai tinggi. Sementara itu, nilai  $R^2$  digunakan untuk menunjukkan kemampuan model dalam menjelaskan variasi data *pace* lari berdasarkan variabel-variabel yang digunakan. Nilai MAE dan MSE yang lebih rendah serta nilai  $R^2$  yang mendekati 1 menunjukkan bahwa model memiliki kinerja prediksi yang baik dan sesuai dengan tujuan analisis.

### 2.6 Deployment

Tahap deployment, Pada tahap ini, model *Random Forest Regressor* yang telah melalui proses pelatihan dan evaluasi digunakan untuk memberikan informasi dan rekomendasi terkait *pace* lari mahasiswa Universitas Darussalam.

Hasil dari analisis feature importance dimanfaatkan untuk memberikan pemahaman mengenai faktor-faktor dominan yang memengaruhi *pace* lari. Implementasi hasil penelitian ini diharapkan dapat menjadi dasar dalam penyusunan program latihan yang lebih terarah dan berbasis data, baik bagi mahasiswa maupun pihak terkait, seperti pelatih atau pengelola kegiatan olahraga kampus. Dengan demikian, tahap *deployment* tidak hanya berfokus pada penerapan teknis model, tetapi juga pada pemanfaatan hasil analisis sebagai alat pendukung pengambilan keputusan dalam upaya peningkatan performa lari mahasiswa.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari hasil pengisian kuesioner oleh mahasiswa Universitas Darussalam Gontor yang dikumpulkan selama periode enam bulan. Total data yang diperoleh berjumlah 118 data. Contoh sebagian data yang digunakan dalam penelitian ini disajikan pada Gambar 2.

semester	usia	pace lari per (km)	berat badan	tinggi badan	jumlah latihan lari per minggu
igmail.c	1	18	7	60	163
@jedu	1	18	10	87	171
igmail.o	1	18	3	63kg	163
sl.com	1	19	15	54	165
aghamx	1	19 5 km/ menit		78 176 cm	Setiap hari
igmail.o	1	18	10	49	159
idqat@x	1	18 2 menit		58	171
mail.com	1	19 10 menit		56	165
gmail.co 1 🍌🍌🍌		21	15	1	168
l.com	1	18	3	56	167
igmail.o	1	17 1,5		68	168
gmail.co	1	18	5	65	162
igmail.i	1	5 km	46 kg	160 cm	
n001@gg 01		19	10	58	170
gmail.co Semester 1		19 10 km	55-60	170-180	
5677@gg	1	19 15 km	67 kg	174 cm	Tergantung kondisi &
igmail.	1	20	10		
mail.com	1	18	18		

Gambar 2. Hasil Pengumpulan Data

### 3.2 Pembersihan Data

Pada tahap ini dilakukan proses pembersihan data pada setiap kolom. Data yang awalnya tersimpan dalam bentuk string dibersihkan dengan menghapus keterangan yang tidak relevan, seperti penulisan tahun, semester, satuan (kg dan km), tanda hubung, tanda kali, emotikon, serta nilai desimal yang tidak diperlukan. Selain itu, keberadaan nilai kosong (missing values) yang berpotensi menghambat proses analisis selanjutnya juga ditangani pada tahap ini. Proses pembersihan data dilakukan untuk memastikan setiap kolom memiliki format yang lebih sederhana, konsisten, dan siap untuk diolah pada tahap analisis dan pemodelan berikutnya. Conroh sebagian data yang telah melewati proses pembersihan data ada pada Gambar 3.

semester	usia	pace	berat	tinggi	jumlah latihan per minggu
gmail.co	22	8	68	162	3
@jedu	23	4	66	158	1
com	20	5	49	150	1
igmail	21	6.4	68	160	4
igmail	18	7	60	163	4
idqat@x	18	10	87	171	5
igmail.i	18	3	62	163	4
igmail.i	19	11	54	165	4
igmail.i	19	5	78	176	7
igmail.i	18	10	49	159	1
idqat@x	18	3	58	171	3
mail.com	19	10	56	165	3
gmail.co	21	15	50	168	3
l.com	18	3	56	167	4
igmail.i	17	15	68	168	2
gmail.co	18	5	65	162	2
igmail	18	5	46	160	5
n001@gg	19	10	58	170	2
gmail.co	19	10	55	170	3
n001@gg	19	15	67	174	3
igmail	20	10	50	162	3
mail.com	19	10	46	167	3
igmail	18	18	60	171	3

Gambar 3. Hasil Pembersihan Data

### 3.3 Transformasi data

Perubahan tipe data pada setiap kolom agar sesuai dengan kebutuhan analisis dan pemodelan. Data yang semula bertipe string dikonversi menjadi tipe numerik, yaitu integer float dan boolean. Variabel semester, usia, tinggi badan, berat badan, dan frekuensi latihan per minggu dikonversi ke dalam tipe data integer. Variabel status perokok direpresentasikan dalam bentuk boolean, sedangkan variabel pace dan indeks massa tubuh dikonversi ke dalam tipe data float. Proses konversi tipe data ini bertujuan untuk memastikan seluruh variabel dapat diproses secara optimal oleh algoritma pembelajaran mesin pada tahap pemodelan. Hasil dari dari tranformasi data dapat dilihat pada Gambar 4.

semester	int64
usia	int64
berat	int64
tinggi	int64
latihan_per_minggu	int64
imt	float64
perokok	bool
pace	float64

Gambar 4. Hasil Transformasi Data

### 3.4 Augmentasi data

Satu data asli direplikasi menjadi lima data baru. Augmentasi dilakukan dengan menambahkan variasi acak kecil pada variabel pace lari menggunakan distribusi uniform dalam rentang 0 hingga 0,3, sehingga nilai pace hasil augmentasi tetap berada dalam batas yang realistis dan mencerminkan fluktuasi alami performa lari mahasiswa. Sebagai contoh, satu data asli dengan nilai pace sebesar 6,20 menit/km setelah proses augmentasi menghasilkan

beberapa variasi nilai pace, antara lain 6,23 menit/km, 6,31 menit/km, 6,38 menit/km, 6,41 menit/km, dan 6,47 menit/km, sementara nilai variabel independen lainnya tetap sama. Perbandingan antara data asli dan data hasil augmentasi, sebagaimana ditunjukkan pada Gambar 5.

Data asli						Data augmentasi					
Email	semester	usia	pace	berat	tinggi	Email	semester	usia	pace	berat	tinggi
ai.com	7	22	8.0	68		ai.com	7	22	8.26	68	
						ai.com	7	22	8.04	68	
						ai.com	7	22	8.03	68	
						ai.com	7	22	8.05	68	
						ai.com	7	22	8.12	68	

Gambar 5. Hasil Augmentasi

3.5 Modeling

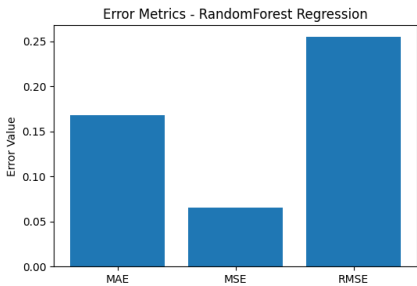
Pada tahap pelatihan, model dikonstruksi menggunakan sebanyak 552 data sebagai data latih, sementara 138 data lainnya dialokasikan sebagai data uji. Proses pelatihan dilakukan dengan menerapkan algoritma Random Forest yang menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan robust. Selama proses tersebut, Model Random Forest Regressor dibangun dengan menggunakan 200 pohon keputusan ( $n\_estimators = 200$ ) dan nilai  $random\_state = 42$  untuk menjaga konsistensi hasil pelatihan. Detail konfigurasi model bisa dilihat pada Tabel 1.

Tabel 1. Konfigurasi Pelatihan Model *Random Forest*

Parameter	Jumlah
Data latih	552
Data uji	138
Pohon keputusan	200

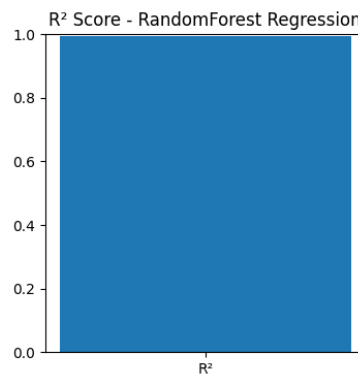
3.6 Evaluasi

Hasil pengujian model Random Forest Regressor menunjukkan kinerja yang sangat baik dalam memprediksi nilai pace. Nilai Mean Absolute Error (MAE) sebesar 0,16 mengindikasikan bahwa rata-rata selisih absolut antara nilai prediksi dan nilai aktual relatif kecil. Nilai Mean Squared Error (MSE) sebesar 0,06 serta Root Mean Squared Error (RMSE) sebesar 0,25 menunjukkan bahwa kesalahan prediksi yang dihasilkan model berada pada tingkat yang rendah, sehingga model mampu menghasilkan prediksi yang cukup akurat dan stabil. Visualisasi error pada *Random forest regressor* ada pada Gambar 6.



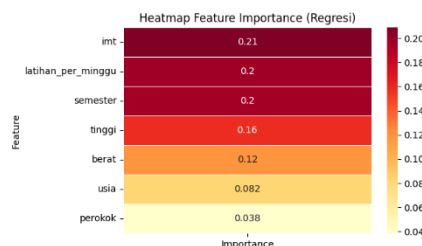
Gambar 6. Metriks *Error* Pada Model

Selain itu, nilai koefisien determinasi ( $R^2$ ) sebesar 0,99 menunjukkan bahwa sebesar 99% variasi nilai pace dapat dijelaskan oleh variabel-variabel yang digunakan dalam model. Nilai  $R^2$  yang mendekati 1 menandakan bahwa model memiliki kemampuan yang sangat baik dalam merepresentasikan hubungan antara variabel input dan variabel target. Secara keseluruhan, hasil ini menunjukkan bahwa model Random Forest Regressor memiliki tingkat akurasi dan kemampuan prediksi yang sangat tinggi. Ringkasan hasil evaluasi kinerja model *Random Forest Regressor* ditunjukkan pada Gambar 7.



Gambar 7. Koefisien Determine

### 3.7 Deployment



Gambar 8. Heatmap Feature Importance

Hasil feature importance untuk menganalisis faktor-faktor yang paling berpengaruh terhadap kecepatan lari menggunakan model regresi, sebagaimana ditunjukkan pada Gambar 8, menunjukkan bahwa Indeks Massa Tubuh (IMT) merupakan faktor yang paling dominan dalam memengaruhi pace lari mahasiswa. Selanjutnya, faktor yang memiliki pengaruh besar adalah frekuensi latihan per minggu dan semester, yang keduanya menunjukkan tingkat kepentingan yang hampir sama. Faktor tinggi badan berada pada urutan berikutnya dengan kontribusi sedang, diikuti oleh berat badan dan usia yang memiliki pengaruh relatif lebih kecil. Adapun status perokok merupakan faktor dengan tingkat kepentingan paling rendah dalam memengaruhi pace lari pada penelitian ini.

Dari hasil ini membuktikan index massa tubuh atau peningkatan massa lemak tubuh memang menyebabkan penurunan kecepatan lari [14]. Secara fisiologis, tubuh dengan massa berlebih terutama lemak berlebih akan meningkatkan beban yang harus digerakkan otot pada setiap langkah lari, sehingga mempengaruhi efisiensi dan kecepatan. Hal ini didukung oleh studi teoretis bahwa performa lari memiliki dasar struktural: massa tubuh dan distribusi otot/lemaknya mempengaruhi gaya dukung dan efisiensi biomekanik saat berlari [15].

## 4. KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis faktor-faktor yang memengaruhi pace lari mahasiswa menggunakan model Random Forest Regressor, serta evaluasi feature importance menggunakan metode Mean Decrease Impurity (MDI), diperoleh beberapa kesimpulan sebagai berikut:

1. Indeks Massa Tubuh (IMT) merupakan faktor yang paling berpengaruh terhadap pace lari, dengan nilai feature importance sebesar 0,21. Hal ini menunjukkan bahwa perubahan IMT memberikan kontribusi terbesar dalam memengaruhi kecepatan lari mahasiswa.
2. Frekuensi latihan per minggu menempati posisi berikutnya dengan nilai 0,20. Hasil ini menunjukkan bahwa konsistensi latihan memiliki peran yang sangat penting dalam meningkatkan performa lari. Mahasiswa yang berlatih lebih sering cenderung memiliki daya tahan dan efisiensi gerak yang lebih baik sehingga mampu mencapai pace lari yang lebih optimal.
3. Semester juga memiliki tingkat kepentingan yang sama besar, yaitu 0,20. Hal ini mengindikasikan bahwa perbedaan jenjang semester berpotensi memengaruhi pace lari mahasiswa, yang dapat dikaitkan dengan perubahan gaya hidup, tingkat kebugaran, pengalaman aktivitas fisik, serta beban akademik yang berbeda pada setiap tingkat pendidikan.
4. Tinggi badan memiliki nilai feature importance sebesar 0,16, yang menunjukkan bahwa aspek antropometri berperan cukup signifikan terhadap kecepatan lari. Tinggi badan berkaitan dengan panjang tungkai dan pola langkah, yang dapat memengaruhi efisiensi gerakan saat berlari.

5. Berat badan memberikan pengaruh sedang dengan nilai 0,12. Kontribusi ini lebih rendah dibandingkan IMT karena sebagian besar informasi terkait komposisi tubuh telah direpresentasikan secara lebih komprehensif oleh variabel IMT.
6. Usia memiliki pengaruh yang relatif kecil dengan nilai 0,082. Hal ini disebabkan oleh rentang usia responden yang tidak terlalu jauh, sehingga variasi usia tidak memberikan dampak yang signifikan terhadap perbedaan pace lari.
7. Status perokok merupakan faktor dengan pengaruh paling rendah, dengan nilai feature importance sebesar 0,038. Hasil ini menunjukkan bahwa dalam populasi mahasiswa yang diteliti, perilaku merokok belum memberikan kontribusi yang signifikan terhadap variasi pace lari, kemungkinan karena intensitas dan proporsi perokok yang relatif kecil.

## 5. SARAN

Secara keseluruhan, model menunjukkan bahwa faktor komposisi tubuh (IMT dan tinggi badan) lebih dominan dalam memengaruhi performa lari dibandingkan faktor kebiasaan, perilaku, maupun karakteristik demografis. Temuan ini sesuai dengan literatur yang menyatakan bahwa antropometri memiliki hubungan kuat dengan performa fisik dan efisiensi gerak.

1. Penelitian selanjutnya disarankan untuk memasukkan parameter fisiologis tambahan seperti  $VO_2$  max, detak jantung saat latihan, atau komposisi lemak tubuh untuk memperoleh gambaran yang lebih komprehensif mengenai faktor penentu kecepatan lari.
2. Dataset dapat diperluas baik dari sisi jumlah responden maupun keberagaman latar belakang aktivitas fisik agar model lebih general dan representatif.
3. Penggunaan teknik interpretasi model tambahan, seperti Permutation Importance atau SHAP values, dapat dilakukan untuk melengkapi analisis dan meminimalkan bias yang mungkin muncul pada metode MDI.

## DAFTAR PUSTAKA

- [1] N. Seethapathi and M. Srinivasan, "Step-to-step variations in human running reveal how humans run without falling," pp. 1–24, 2019.
- [2] M. I. Hafizhuddin, "Hubungan Antara Self Disclosure Melalui Status Wa Dan Kualitas Hidup Pada Mahasiswa Di Universitas Muhammadiyah Surabaya," *Skrip*, p. h. 2, 2019, [Online]. Available: [http://repository.um-surabaya.ac.id/id/eprint/3715%0Ahttp://repository.um-surabaya.ac.id/3715/3/BAB\\_II.pdf](http://repository.um-surabaya.ac.id/id/eprint/3715%0Ahttp://repository.um-surabaya.ac.id/3715/3/BAB_II.pdf)
- [3] Eraspaces.com, "mengenal-apa-itu-pace-dalam-lari-untuk-optimalkan-latihanmu @ eraspaces.com," eraspaces.com. [Online]. Available: <https://eraspaces.com/artikel/post/mengenal-apa-itu-pace-dalam-lari-untuk-optimalkan-latihanmu>
- [4] I. M. Y. Parwata, "Hubungan Tinggi Badan dan Berat Badan Terhadap Kecepatan Lari 100 Meter Mahasiswa Putra Fpik Ikip PGRI Bali," *J. Pendidik. Kesehat. Rekreasi*, vol. 3, no. 2, pp. 19–27, 2017, doi: 10.59672/jpkr.v3i2.232.
- [5] G. D. Padaningpuri and E. B. P. S. Andar, "Hubungan Faktor-Faktor Kinesiologi dengan Kecepatan Lari 100 Meter Mahasiswa Fk Undip," vol. 8, no. 2, pp. 713–722, 2019.
- [6] L. H. Quoc, "Current State of Factors Influencing the Development of Running Speed in the 100m Sprint for Non-Specialized University Students at Tan Trao University," *Eur. J. Arts, Humanit. Soc. Sci.*, vol. 1, no. 3, pp. 335–345, May 2024, doi: 10.59324/ejahss.2024.1(3).29.
- [7] M. S. Hasan and A. Sukanto, "Kontribusi Kecepatan Reaksi Kaki dan Kelentukan Terhadap Kemampuan Lari 100 meter pada Mahasiswa PKO FIK UNM Contribution of Foot Reaction Speed and Flexibility to the Ability to Run 100 meters in PKO FIK UNM Students," vol. 1, no. 1, pp. 14–21, 2021.
- [8] I. Ikadarny and A. Karim, "Kontribusi Kecepatan Lari, Daya Ledak Tungkai, Dan Keseimbangan Terhadap Kemampuan Lompat Jauh Murid SMP Negeri 2 Kabupaten Gowa," *Jendela Olahraga*, vol. 6, no. 1, pp. 158–164, 2021, doi: 10.26877/jo.v6i1.6971.
- [9] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-streib, "Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing," vol. 4, pp. 1–20, 2021, doi: 10.3389/frai.2021.576892.
- [10] D. Muriyatmoko, D. Fikrianti, and F. R. Ifalus, "Analisis Penjualan Produk Terlaris di Toko Bangunan Pekanbaru Jaya Menggunakan Metode Clustering K-Means," vol. 4, pp. 88–93, 2025.
- [11] A. Y. Firmansyah, "Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja," vol. 13, no. 1, Kediri, 24 Januari 2026

pp. 397–407, 2024.

- [12] D. Rawan, T. Pidana, and N. Di, “Implementasi K-Means Clustering dan Model Crisp-Dm Untuk Pengelompokan Daerah Rawan Tindak Pidana Narkoba Di Diy,” vol. 7, no. 2, pp. 606–615, 2025.
- [13] H. A. Damayanti and U. P. Sanjaya, “Perbandingan Model Pembelajaran Mesin Berbasis Smote Meningkatkan Identifikasi Siswa Berisiko di Sekolah Menengah Pertama,” vol. 12, no. 1, pp. 119–127, 2025, doi: 10.30656/jsii.v11i2.9065.
- [14] L. Genton *et al.*, “An Increase in Fat Mass Index Predicts a Deterioration of Running Speed.,” *Nutrients*, vol. 11, no. 3, Mar. 2019, doi: 10.3390/nu11030701.
- [15] P. G. Weyand and J. A. Davis, “Running performance has a structural basis,” pp. 2625–2631, 2005, doi: 10.1242/jeb.01609.