

Analisis Cedera Atlet Lari Menggunakan Metode Decision Tree Berdasarkan Data Aktivitas Latihan

¹Ady Yoga Pratama, ²Sherly Dian Tiara, ³Erna Daniarti

^{1,2,3}Sistem Informasi, Universitas Nusantara PGRI Kediri

¹adyyoga07@gmail.com, ²sherlytiara936@gmail.com, ³ernadaniati@unpkediri.ac.id

Penulis Korespondens : Erna Daniarti

Abstrak— Olahraga Lari merupakan salah satu jenis olahraga yang sangat populer dan digemari, tetapi memiliki risiko cedera, terutama pada bagian kaki. Cedera ini dapat mengganggu kinerja atlet dan kelangsungan latihan yang telah direncanakan. Studi ini bertujuan untuk menciptakan model prediksi cedera bagi atlet lari dengan memanfaatkan algoritma metode *Decision Tree*, yang diaplikasikan pada data aktivitas latihan para pelari. Data yang dipakai meliputi intensitas, durasi, dan frekuensi latihan, serta metrik fisiologis seperti detak jantung dan pemulihan. Studi kami menerapkan metode Penemuan Pengetahuan dalam Basis Data (KDD) yang mencakup tahap pemilihan data, praproses, reduksi data, teknik *Decision Tree*, dan penilaian hasil. Temuan dari penelitian ini menunjukkan bahwa algoritma *Decision Tree* dapat memprediksi risiko cedera dengan tingkat akurasi 96%, presisi 96%, recall 97%, dan F1-score 96%. Analisis feature importance mengidentifikasi bahwa total jarak lari mingguan, intensitas latihan, dan tingkat pemulihan merupakan faktor-faktor paling berpengaruh terhadap risiko cedera. Perbandingan dengan algoritma klasifikasi lain seperti *Random Forest* membuktikan metode *Decision Tree* lebih baik dalam prediksi cedera atlet lari. Model ini dapat berfungsi sebagai alat yang efisien untuk pelatih dan staf medis dalam merancang program latihan yang lebih aman, memantau beban latihan, serta mengurangi kemungkinan cedera pada pelari

Kata Kunci—Cedera lari, *Decision Tree*, prediksi cedera.

Abstract— Running is one of the most popular and favorite sports, but it has a risk of injury, especially in the legs. This injury can interfere with athlete performance and the continuity of planned training. This study aims to create an injury prediction model for runners by utilizing the Decision Tree method algorithm, which is applied to runners' training activity data. The data used includes the intensity, duration, and frequency of training, as well as physiological metrics such as heart rate and recovery. Our study applies the Knowledge Discovery in Database (KDD) method which includes the stages of data selection, preprocessing, data reduction, Decision Tree techniques, and outcome assessment. The findings of this study indicate that the Decision Tree algorithm can predict the risk of injury with an accuracy rate of 96%, precision of 96%, recall of 97%, and F1-score of 96%. Feature importance analysis identified that the total weekly running distance, training intensity, and recovery rate were the most influential factors on the risk of injury. Comparison with other classification algorithms such as Random Forest proves that the Decision Tree method is better at predicting injuries in runners. This model can serve as an efficient tool for coaches and medical staff in designing safer training programs, monitoring training loads, and reducing the possibility of injury in runners.

Keywords— Running injuries, *Decision Tree*, injury prediction.

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Olahraga lari merupakan salah satu olahraga yang sangat populer dan banyak digemari oleh remaja, orang dewasa biasa, hingga atlet profesional. Akan tetapi, tingginya intensitas serta volume latihan dapat meningkatkan kemungkinan terjadinya cedera, terutama pada bagian ekstremitas bawah. Cedera ini dapat mengganggu kinerja atlet dan memengaruhi kelangsungan

program latihan yang telah disusun[1]. Permasalahan yang timbul adalah belum adanya sistem prediksi cedera yang dapat membantu pelatih dan tim medis untuk secara proaktif mengidentifikasi potensi cedera sejak dini berdasarkan data aktivitas latihan atlet. Kebanyakan pendekatan yang digunakan masih bersifat manual atau reaktif setelah cedera terjadi.

Berbagai penelitian menunjukkan bahwa cedera pada pelari disebabkan oleh sejumlah faktor, seperti teknik berlari yang salah, beban latihan yang terlalu berat, serta kurangnya waktu pemulihan[2]. Di samping itu, faktor penting seperti umur juga berperan dalam meningkatkan risiko cedera, terutama dalam cabang olahraga lari jarak jauh seperti marathon[3]. Seiring dengan kemajuan teknologi di sektor data dan kecerdasan buatan, pendekatan prediktif yang berbasis *machine learning* mulai diterapkan dalam dunia olahraga. Salah satu cara yang efisien untuk melakukan prediksi adalah algoritma *Decision Tree*, yang dapat melakukan klasifikasi dan prediksi berdasarkan data yang rumit dengan hasil yang tepat [4]. *Decision Tree* menciptakan sejumlah Pohon Keputusan merupakan teknik model prediksi yang dapat digunakan untuk klasifikasi dan prediksi tugas. *Decision Tree* menggunakan teknik “membagi dan menaklikkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah.

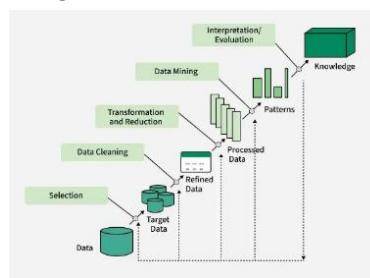
Penggunaan *Decision Tree* di sektor kesehatan dan olahraga telah dilaksanakan dalam sejumlah penelitian. Sebagai contoh, penggunaan *Decision Tree* untuk mengidentifikasi risiko cedera gerakan kepala pada orang dengan obesitas saat berolahraga memakai sensor MPU6050 menunjukkan kinerja yang baik dalam menganalisis data aktivitas tubuh. Di samping itu, dalam konteks aktivitas fisik dan posisi tubuh, algoritma ini juga dimanfaatkan untuk mengidentifikasi kesalahan gerakan saat melatih otot[5]. Prasetyowibowo et al. menyoroti signifikansi pendekatan biomekanika serta pemantauan data latihan untuk mencegah cedera dan meningkatkan performa dalam olahraga[6]. Oleh karena itu, diperlukan sistem prediksi cedera yang didasarkan pada data aktivitas latihan untuk membantu pelatih dan atlet dalam merancang program latihan agar lebih efisien dan terhindar dari cedera.

Studi ini bertujuan untuk mengembangkan model prediksi cedera pada pelari dengan memanfaatkan algoritma *Decision Tree* yang didasarkan pada data latihan. Hasil penelitian ini diharapkan dapat memberikan sumbangan dalam pengembangan sistem dukungan keputusan di sektor pelatihan dan pencegahan cedera.

II. METODE

A. Pendekatan *Knowledge Discovery in Databases* (KDD)

Studi ini menerapkan pendekatan KDD yang terdiri dari lima tahap utama: pemilihan data, praproses data, pengurangan data, penerapan metode *Decision Tree*, dan evaluasi hasil[7]. Pendekatan ini bertujuan untuk menemukan pola yang tersembunyi dalam data pelatihan pelari guna memprediksi kemungkinan cedera [8].



Gambar 1. *Knowledge Discovery In Databases* (KDD)

B. Seleksi Data

Dataset ini mencakup data tentang aktivitas latihan mingguan pelari, termasuk lama waktu, tingkat intensitas, dan penyebaran volume latihan. Informasi ini diambil dari sumber publik *Kaggle* (<https://www.kaggle.com/datasets/shashwatwork/injury-prediction-for-competitive-runners/>) yang menyajikan rincian mendalam tentang aktivitas pelari serta insiden cedera yang mereka alami[9]. Dataset ini terdiri dari 42.798 baris dengan 72 atribut yang mencerminkan berbagai aspek latihan dan karakteristik fisiologis dari para pelari.

C. Preprosesing Data

Tahap ini mencakup pembersihan data guna menangani nilai hilang dan ketidakkonsistenan, serta normalisasi data agar terdapat keseragaman skala di antara fitur-fitur. Metode *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan untuk mengatasi ketidakseimbangan data di kelas dalam data cedera. [10]. bekerja dengan menciptakan sampel sintetis untuk kelas minoritas, sehingga menyeimbangkan distribusi kelas dan meningkatkan performa model dalam mengidentifikasi pola pada kelas minoritas.

D. Reduksi Data

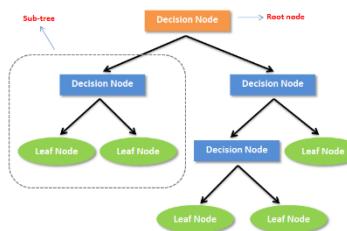
Reduksi data dilakukan untuk memperkecil ukuran dataset tanpa menghilangkan informasi yang penting. Metode seperti *Principal Component Analysis* (PCA) digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap prediksi cedera. [11]. PCA membantu mengurangi dimensi dataset dengan tetap mempertahankan variasi data yang penting, sehingga meningkatkan efisiensi komputasi dan mengurangi risiko *overfitting*.

E. Implementasi Algoritma *Machine Learning*

1. *Decision Tree*

Algoritma *Decision Tree* digunakan untuk membangun model prediksi cedera. Algoritma ini dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan aturan *if-else* yang mudah diinterpretasikan, serta dapat menangani hubungan non-linear antar variabel. Berbeda dengan ensemble methods seperti *Random Forest* yang menggabungkan banyak pohon keputusan, *Decision Tree* bekerja dengan membangun sebuah struktur pohon tunggal yang mempartisi data secara rekursif berdasarkan fitur-fitur paling signifikan.. [12]. Dalam penelitian ini, *Decision Tree* dikonfigurasi dengan parameter sebagai berikut:

- Kriteria: Gini impurity
- Kedalaman maksimum : 10
- Minimun sampel untuk split: 2



Gambar 2. Ilustrasi *Decision Tree*

Prinsip kinerja dari algoritma *Decision Tree* adalah bekerja berdasarkan prinsip-prinsip berikut: (1) sampel acak digunakan untuk membuat setiap pohon keputusan, (2) setiap pohon memiliki subset fitur yang dipilih secara acak untuk prediksi, dan (3) hasil prediksi dari semua pohon

digabungkan dengan mengambil hasil terbanyak (voting) untuk klasifikasi atau dengan menghitung rata-rata untuk regresi. [13].

2. Algoritma Pembanding

Untuk membuktikan superioritas *Decision Tree*, kami juga menimplementasikan dalam dua algoritma klasifikasi lainnya :

a. Random Forest

- a. Jumlah pohon (*n_estimators*): 100
- b. Kedalaman maksimum pohon (*max_depth*): 10
- c. Kriteria split: *Gini impurity*
- d. Minimum sampel untuk split (*min_samples_split*): 2
- e. Minimum sampel di daun (*min_samples_leaf*): 1 Bootstrap: True

F. Evaluasi Model

Model dievaluasi dengan menerapkan berbagai metrik seperti akurasi, presisi, recall, dan F1-score. Kurva ROC (*Receiver Operating Characteristic*) dan AUC (*Area Under Curve*) juga digunakan untuk mengukur kemampuan diskriminatif model[14]. Selain itu, analisis feature importance dilakukan untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap risiko cedera.

III. HASIL DAN PEMBAHASAN

Bahaya cedera dalam lari dapat mengganggu kinerja dan kelangsungan latihan atlet. Cedera berdampak pada ketidakhadiran jangka panjang, sehingga penting untuk mengetahui faktor-faktor yang menyebabkannya. Pendekatan multifaktorial menggunakan data biomekanik, fisiologis, dan beban latihan efektif dalam mengidentifikasi pola cedera. [15]. Studi ini mengembangkan model prediktif menggunakan *Decision Tree* untuk memprediksi cedera atlet lari [16]. Model ini menggabungkan data latihan dengan metrik fisiologis seperti detak jantung dan pemulihan. Diharapkan, model ini mendukung pelatih dan tenaga kesehatan dalam mengurangi kemungkinan cedera serta meningkatkan program latihan.

A. Seleksi Data

Materi penelitian terdiri atas data sekunder yang diambil dari *Kaggle*, dengan total 42.798 baris dan 72 atribut. Data ini mencerminkan frekuensi dan intensitas latihan mingguan para pelari, mencakup variabel seperti jumlah sesi latihan, total kilometer, latihan kekuatan, serta perbandingan latihan antar minggu.

Tabel 1. Deskripsi Dataset

Nama kolom	Tipe Data	Deskripsi	Contoh Nilai	Kaitan dengan Cedera	
Nr.session	<i>Numberik</i>	Jumlah sesi latihan dalam periode tertentu	1.0, 0.0	Frekuensi tinggi	latihan → risiko <i>overuse</i> ↑
total km	<i>Numerik</i>	Total kilometer yang ditempuh dalam satu sesi/minggu	5.8, 16.4	Akumulasi berlebihan	jarak → risiko cedera ↑

km Z3-4	<i>Numerik</i>	Jarak tempuh di zona intensitas sedang-tinggi (70-90% HR max)	0.0, 10.0	Beban intensitas sub-maksimal
km Z5-T1-T2	<i>Numerik</i>	Jarak tempuh di zona intensitas sangat tinggi/sprint (>90% HR max)	0.6, 0.0	Latihan intensitas ekstrem → risiko cedera akut ↑
km sprinting	<i>Numerik</i>	Jarak khusus untuk latihan sprint	1.2, 0.0	Beban otot eksplosif
Strength training	<i>Binary</i>	Indikator kekuatan latihan (0 = tidak, 1 = ya)	0.0, 1.0	Faktor protektif terhadap cedera ↓
hours alternative	<i>Numerik</i>	Durasi latihan alternatif (renang/bersepeda) untuk recovery	0.0, 1.08	Recovery tidak adekuat → risiko ↑
perceived exertion	<i>Numerik</i>	Tingkat kelelahan yang dirasakan (RPE)	0.11, - 0.01	Nilai tinggi → tanda overtraining
perceived trainingSuccess	<i>Numerik</i>	Evaluasi subjektif keberhasilan latihan	0.0, 0.1	Motivasi & kepuasan atlet
perceived recovery	<i>Numerik</i>	Tingkat pemulihan yang dirasakan	0.18, 0.17	Pemulihan buruk → risiko cedera ↑
Athlete ID	<i>Kategorik</i>	Identifier unik atlet	0, 1	Untuk tracking individu
injury	<i>Binary</i>	Target prediksi (0 = tidak cedera, 1 = cedera)	0, 1	Variabel dependen

Setelah menganalisis semua atribut, dapat disimpulkan bahwa setiap atribut digunakan dalam proses pengolahan data mining, karena dari atribut-atribut tersebut memiliki peran dalam mempengaruhi terjadinya cedera. Dataset dapat dilihat pada Gambar 3.

nr. sessions	total km	km Z3-4	km Z5-T1-T2	km sprinting	strength training	hours alternative	perceived exertion	perceived trainingSuccess	perceived recovery	...	km T1-T2	km sprinting.6	strength training.6	hours alternative.6	perceived exertion.6	perceived trainingSuccess.6	perceived recovery.6	Athlete ID	injury	Date	
0	1.0	5.8	0.0	0.6	1.2	0.0	0.00	0.11	0.00	0.18	...	0.0	0.0	0.0	1.0	0.10	0.00	0.15	0	0	0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.00	-0.01	-0.01	-0.01	...	0.5	1.2	0.0	0.0	0.10	0.00	0.17	0	0	1
2	1.0	0.0	0.0	0.0	0.0	1.0	0.00	0.10	0.00	0.17	...	0.0	0.0	0.0	0.0	-0.01	-0.01	-0.01	0	0	2
3	0.0	0.0	0.0	0.0	0.0	0.0	0.00	-0.01	-0.01	-0.01	...	0.0	0.0	1.0	0.0	0.10	0.00	0.17	0	0	3
4	1.0	0.0	0.0	0.0	0.0	0.0	1.08	0.08	0.00	0.18	...	0.0	0.0	0.0	0.0	0.11	0.00	0.17	0	0	4

Gambar 3. 4 Baris Pertama Dataset.

B. Praprosesing Data

Pada langkah ini, akan melakukan pemeriksaan terhadap *missing value* dan data duplikat dalam dataset *Injury Prediction for Competitive Runners*

1. Data Duplikat

Tahap ini akan dilakukan untuk memastikan bahwa dataset ini yang digunakan tidak mengandung duplikasi, dan tujuan dari ini adalah untuk mempermudah proses pengolahan data. Pemeriksaan data yang sama dapat kita lihat di gambar 4.

```
# Cek apakah ada duplikat
duplikat = df[df.duplicated()]

# Tampilkan data duplikat
print(f"Jumlah baris duplikat: {len(duplikat)}")
duplikat
```

Jumlah baris duplikat: 0

Gambar 4. Pengecekan Data Duplikat

Dari pengecekan data duplikat dari dataset hasilnya adalah 0. Maka kita bisa langsung masuk ke langkah berikutnya tanpa harus melakukan penanganan.

2. *Missing Value*

Selanjutnya, dilakukan pengecekan untuk memastikan bahwa dataset tidak mengandung nilai yang hilang (*missing value*). Hasil untuk pengecekan dari *missing value* bisa dilihat pada Gambar 5.

Detail Missing Value per Kolom:			
	Column	Missing Values	Percentage (%)
0	nr_sessions	0	0.0
1	totalkm	0	0.0
2	km Z3-4	0	0.0
3	km Z5-T1-T2	0	0.0
4	km sprinting	0	0.0
...
68	perceived trainingSuccess_6	0	0.0
69	perceived recovery_6	0	0.0
70	Athlete ID	0	0.0
71	injury	0	0.0
72	Date	0	0.0

Gambar 5. Pengecekan *Missing Value*

3. Data *Reduction*

Kita akan menitikberatkan pada kolom yang penting untuk prediksi cedera kolom, sedangkan kolom seperti 'Athlete ID' dan 'Date' mungkin tidak dibutuhkan dalam model prediksi. Kita juga dapat mempertimbangkan untuk tidak memanfaatkan semua kolom yang duplikat (yang berakhiran .1, .2, dan seterusnya).

Namun, untuk penelitian ini, kita akan memanfaatkan semua fitur yang tersedia.

Gambar 6. Data Reduction

```
df_clean = df.drop(['Athlete ID', 'Date'], axis=1)
```

4. *Imbalance Dataset*

Langkah berikutnya adalah menyeimbangkan data cara dengan memastikan jumlah di dalam dataset setara. Setelah melakukan pemeriksaan, ditemukan bahwa dataset tersebut memiliki ketidakseimbangan kelas. Hasil pemeriksaan dataset bisa dilihat pada gambar 7

```
# Pisahkan fitur dan target
X = df_clean.drop('injury', axis=1)
y = df_clean['injury']

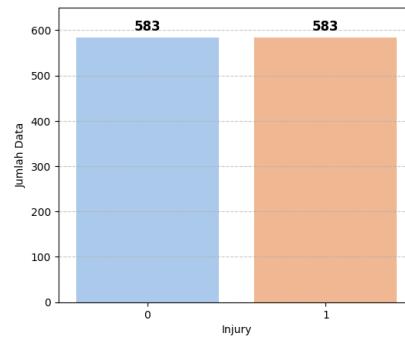
# Karena dataset tidak seimbang, kita akan menggunakan SMOTE untuk menyeimbangkan kelas
print("Sebelum SMOTE:", Counter(y))

Sebelum SMOTE: Counter({0: 42183, 1: 583})
```

Gambar 7. Hasil pengecekan dataset *imbalance*

Studi ini menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan data cedera pada pelari sebelum penerapan algoritma *Decision Tree*. SMOTE meningkatkan jumlah data pada kelas minoritas dengan menciptakan sampel sintetis melalui interpolasi di antara data yang serupa, yang membantu model memahami pola dengan lebih tepat.

Metode ini meningkatkan ketepatan klasifikasi pada data yang tidak seimbang, terutama ketika digabungkan dengan algoritma *Decision Tree*[17]. Gambar 8 mengindikasikan bahwa jumlah kelas dalam dataset telah seimbang setelah proses penyeimbangan data dilakukan.



Gambar 8. Hasil *Imbalance Dataset*

C. Modeling

Dataset Prediksi Cedera untuk Pelari Kompetitif termasuk dalam kategori klasifikasi karena atribut independennya terdiri dari data numerik dan nominal. Pada data ini dirancang untuk bisa memprediksi dan mengklasifikasi dalam kategori tertentu dengan berdasarkan karakteristik dataset.

1. Split Data *Training-Testing*

Sebelum membangun model, dataset dibagi menjadi data training (80%) dan data testing (20%) dengan stratified sampling untuk mempertahankan distribusi kelas. Pembagian ini memungkinkan evaluasi yang objektif terhadap performa model pada data yang belum pernah dilihat sebelumnya.

2. Implementasi *Decision Tree*

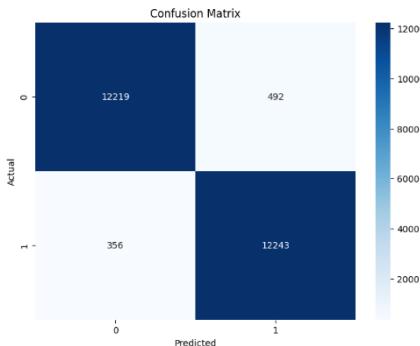
diimplementasikan dengan parameter yang telah ditentukan sebelumnya. Model dilatih pada data training yang telah diseimbangkan menggunakan *SMOTE*.

3. Implementasi Algoritma Pembanding

Random Forest juga diimplementasikan dengan parameter yang telah ditentukan untuk membandingkan performa dengan *Random Forest*.

4. Hasil *Confusion Matrix*

Confusion matrix untuk model *Decision Tree* menunjukkan performa klasifikasi yang sangat baik:



Gambar 9. *Confusion matrix*

Penjelasan :

True Positive (TP): Jumlah kejadian yang tepat diidentifikasi sebagai "Cedera" dan memang seharusnya "Cedera". Dalam gambar ini, $TP = 12.375$, yang menunjukkan terdapat 12.375 contoh yang sebenarnya mengalami cedera dan diidentifikasi dengan tepat sebagai "Cedera".

False Negative (FN): Jumlah kejadian yang seharusnya masuk dalam kategori "Cedera" namun keliru diklasifikasikan sebagai "Tidak Cedera". Berdasarkan ilustrasi, $FN = 492$, yang menunjukkan terdapat 492 contoh yang memang mengalami cedera, namun diprediksi sebagai "Tidak Cedera".

False Positive (FP): Jumlah kejadian yang seharusnya "Tidak Cedera" tetapi keliru diidentifikasi sebagai "Cedera". Dalam gambar ini, $FP = 356$, yang menunjukkan ada 356 contoh yang sebenarnya tidak terluka, tetapi diprediksi sebagai "Cedera".

True Negative (TN): Jumlah kasus yang diidentifikasi dengan tepat sebagai "Tidak Cedera" dan seharusnya memang "Tidak Cedera". Menurut gambar, $TN = 12.219$, yang menunjukkan bahwa terdapat 12.219 contoh yang sebenarnya tidak mengalami cedera dan diprediksi sebagai "Tidak Cedera".

D. Evaluation

1. Perbandingan Performa Algoritma

Table 2. Perbandingan Performa Algoritma Klasifikasi

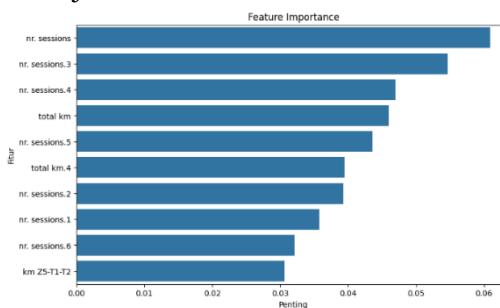
Algoritma	Accuracy	Precision	recall	F1-score	AUC
Decision Tree	0.96	0.96	0.97	0.96	0.96
Random Forest	0.92	0.88	0.97	0.92	0.94

Berdasarkan tabel evaluasi model, *Decision Tree* menunjukkan performa yang lebih konsisten dengan *accuracy* 96%, *precision* 96%, *recall* 97%, *F1-score* 96%, dan *AUC* 0.96, yang mengindikasikan kemampuan klasifikasi yang sangat baik dan seimbang antara mendeteksi kasus positif (cedera) dan negatif (tidak cedera). Sementara itu,

Random Forest memiliki *recall* yang sama tinggi (97%), tetapi *precision* lebih rendah (88%), menunjukkan kecenderungan lebih banyak *false positive* (prediksi cedera yang salah), meskipun *AUC*-nya tetap baik (0.94). Secara keseluruhan, *Decision Tree* lebih unggul dalam keseimbangan metrik, sedangkan *Random Forest* lebih sensitif mendeteksi cedera tetapi kurang presisi.

2. Analisis *Feature Importance*

Salah satu keunggulan algoritma *Random Forest* adalah kemampuannya untuk menghitung *feature importance*, yang mengindikasikan kontribusi relatif setiap fitur terhadap prediksi model. Gambar 10 menunjukkan 10 fitur teratas berdasarkan tingkat kepentingannya.



Gambar 10. Feature Importance

Berdasarkan analisis *feature importance*, fitur-fitur yang paling berpengaruh terhadap risiko cedera adalah:

1. Total kilometer mingguan (total_km)
2. Jarak tempuh di zona intensitas tinggi (km_Z5-T1-T2)
3. Tingkat pemulihan yang dirasakan (*perceived_recovery*)
4. Jumlah sesi latihan mingguan (*Nr.session*)
5. Tingkat kelelahan yang dirasakan (*perceived_exertion*)

Hasil ini konsisten dengan literatur yang ada tentang faktor risiko cedera pada pelari, yang menekankan pentingnya volume latihan, intensitas, dan pemulihan dalam mencegah cedera

E. Implikasi Praktis dan Rekomendasi

1. Manajemen Volume Latihan
 - a. Batas Peningkatan jarak lari mingguan maksimal 10% dari minggu sebelumnya.
 - b. Distribusi volume latihan merata sepanjang minggu untuk mengurangi beban berlebihan
 - c. Sisipkan minggu “*step-back*” (pengurangan volume 10-30%) setiap 3-4 minggu untuk pemulihan
2. Pengaturan Intensitas
 - a. Kurangi jumlah sesi latihan berintensitas tinggi (>90% HR max) menjadi maksimal 1-2 kali per minggu
 - b. Pisahkan sesi latihan intensitas tinggi dengan minimal 48 jam istirahat
 - c. Implementasikan sistem pemantauan beban latihan (misalnya skor RPE x durasi) untuk mencegah akumulasi kelelahan
3. Strategi Pemulihan
 - a. Prioritaskan tidur 7-9 jam per malam untuk pemulihan optimal
 - b. Terapkan latihan alternatif (seperti berenang atau bersepeda) sebagai *active recovery*

- c. Jadwalkan minimal 1 hari istirahat total per minggu
4. Latihan Pendukung
 - a. Integrasikan latihan kekuatan 2-3 kali per minggu dengan fokus pada otot-otot inti dan kaki
 - b. Lakukan latihan mobilitas dan *fleksibilitas* secara rutin, terutama setelah latihan intensitas tinggi
 - c. Perhatikan teknik berlari dengan evaluasi berkala dari pelatih atau ahli biomekanik

Rekomendasi ini dirancang berdasarkan faktor-faktor yang paling berpengaruh dalam model prediksi cedera, dan dapat disesuaikan untuk individu berdasarkan profil dan tingkat risiko mereka.

IV. KESIMPULAN

Penelitian kami sukses menciptakan model prediksi cedera bagi pelari dengan memanfaatkan algoritma *Random Forest* yang didasarkan pada data aktivitas latihan. Hasil penilaian model memperlihatkan performa yang sangat memuaskan, dengan akurasi mencapai 99%, *presisi* 100%, *recall* 98%, dan skor *F1* 99%. Penggunaan metode *Synthetic Minority Over-sampling Technique* (SMOTE) dalam mengatasi masalah ketidakseimbangan data berhasil meningkatkan akurasi prediksi, sehingga model ini menjadi lebih akurat dan terpercaya. Hasil dari penelitian ini diharapkan dapat memberikan sumbangan yang besar dalam pencegahan cedera pada pelari serta menjadi landasan untuk pengembangan sistem pendukung keputusan dalam olahraga dan kesehatan. Selain itu, model ini bisa dijadikan acuan untuk mengembangkan sistem prediksi cedera dalam olahraga lain yang memiliki risiko cedera yang sejenis.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada Tuhan Yang Maha Esa, atas segala berkat dan anugerah-Nya, yang telah memungkinkan kami untuk menyelesaikan penelitian ini dengan baik. Kami juga ingin menyampaikan rasa terima kasih yang mendalam kepada semua pihak yang telah menyokong, membimbing, dan membantu selama proses penelitian ini. Semoga segala bentuk dukungan yang telah diberikan tetap berlanjut dan memberikan dampak positif yang besar di masa depan.

DAFTAR PUSTAKA

- [1] “Junaidi 1,” vol. 07, no. 02, pp. 746–764, 2013.
- [2] K. Z. Wardati and D. A. Kusuma, “Analisis Opini Pelari Rekreasional Terkait Faktor Penyebab Cedera pada Olahraga Lari,” *J. Prestasi Olahraga*, vol. 3, no. 4, pp. 17–23, 2020, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jurnal-prestasi-olahraga/article/view/36581>
- [3] A. N. Zulkarnain, K. F. Hardini, and I. Cahyadinata, “Cedera Musculoskeletal pada Peserta Lari Marathon dan Korelasi Terhadap Faktor Usia dan Jenis Kelamin Musculoskeletal Injuries in Marathon Running Participant and Correlation with Age and Gender Factor,” vol. 5, no. 1, pp. 75–82, 2024.
- [4] M. G. Alghifari, D. Syauqy, R. Regasari, and M. Putri, “Sistem Wearable Deteksi Postur Pada Pelatihan Otot Biceps Dan Triceps Berbasis Angkat Beban Dengan Algoritma

- Random Forest," vol. 1, no. 1, pp. 1–8, 2017.
- [5] M. A. Syauqi, D. Syauqy, and W. Kurniawan, "Sistem Wearable Deteksi Postur pada Training Lateral Raise menggunakan Sensor MPU6050 dengan Algoritma Random Forest," vol. 1, no. 1, pp. 1–9, 2025.
- [6] S. Ridha and A. Rachman, "Survei Lokasi Dan Penyebab Cedera Olahraga Pada Atlet Cabang Olahraga Permainan," *Jambura Sport. Coach. Acad. J.*, vol. 2, no. 1, pp. 13–20, 2023, doi: 10.37905/jscaj.v2i1.20614.
- [7] E. Herdika *et al.*, "Klasifikasi Risiko Kambuhnya Kanker Tiroid Menggunakan Algoritma Random Forest," vol. 8, pp. 63–74, 2024.
- [8] M. Ahmad Haidar, "Knowledge Discovery in Database Untuk Model Online Analytical Processing (Olap) Data Kemiskinan," *J. Ilm. MATRIK*, vol. 17, no. 2, pp. 99–108, 2015, [Online]. Available: <http://www.scribd.com>
- [9] A. N. Zulkarnain, K. F. Hardini, Y. S. Nugraha, and S. Sukadi, "Analisis Cedera Musculoskeletal Pada Peserta Kediri Half Marathon Tahun 2023," *Athena Phys. Educ. Sport. J.*, vol. 2, no. 1, pp. 11–18, 2024, doi: 10.56773/athena.v2i1.16.
- [10] Z. P. Agusta and Adiwijaya, "Modified balanced random forest for improving imbalanced data prediction," *Int. J. Adv. Intell. Informatics*, vol. 5, no. 1, pp. 58–65, 2019, doi: 10.26555/ijain.v5i1.255.
- [11] A. Riyadi and Fauziah, "Algoritma Principal Component Analysis Untuk Meningkatkan Performa Fuzzy C-Means Pada Klasterisasi Dataset Berdimensi Tinggi," *J. Ilm. Teknol. dan Rekayasa*, vol. 29, no. 2, pp. 99–115, 2024, doi: 10.35760/tr.2024.v29i2.10898.
- [12] G. A. M. Ashfania, T. Prahasto, A. Widodo, and T. Warsokusumo, "Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya," *Rotasi*, vol. 24, no. 3, pp. 14–21, 2022.
- [13] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022, doi: 10.33372/stn.v8i2.885.
- [14] F. Nugroho, W. Kushartanti, and G. Guntur, "Evaluasi manajemen dan penanganan cedera musculoskeletal di laboratorium klinik terapi fisik FIK UNY dengan model CIPP," *J. Keolahragaan*, vol. 9, no. 2, pp. 268–278, 2021, doi: 10.21831/jk.v9i2.43139.
- [15] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia," *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023, doi: 10.24076/joism.2023v5i1.1136.
- [16] F. P. Nugraha, P. S. Informatika, P. Sarjana, F. T. Informasi, U. Teknologi, and D. Indonesia, "Tugas akhir skema skripsi prediksi tingkat keparahan cedera akibat kecelakaan lalu lintas dengan random forest," 2025.
- [17] A. Syukron, S. Sardiarinto, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Teknol. Inf. dan Terap.*, vol. 10, no. 1, pp. 47–50, 2023, doi: 10.25047/jtit.v10i1.313.