

# PENDETEKSI PENYAKIT HEPATITIS MENGGUNAKAN CART DECISION TREE

Rosyid Eko Nugroho<sup>1\*</sup>, Wisnu Yogi Pamungkas<sup>2</sup>, Jajam Haerul Jaman<sup>3</sup>

<sup>1,2,3</sup>Universitas Singaperbangsa Karawang, Jl.HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

Received: 12 Agustus 2024  
Accepted: 5 Oktober 2024  
Published: 12 Oktober 2024

## Keywords:

Hepatitis; CART Decision Tree; KDD.

## Correspondent Email:

rosyid123.ren@gmail.com

**Abstrak.** Hepatitis merupakan penyakit menular yang menjadi penyebab utama kematian di seluruh dunia. Penelitian ini mengusulkan penggunaan metode CART decision tree untuk mendeteksi penyakit hepatitis secara dini. Model ini dilatih dan dievaluasi menggunakan data klinis untuk mengidentifikasi pola dan indikator signifikan dalam diagnosis hepatitis. Hasil evaluasi menunjukkan tingkat akurasi yang tinggi sebesar 99,17%, menunjukkan efektivitas model dalam mendeteksi hepatitis. Implementasi model ke dalam sebuah website memungkinkan pengguna untuk menerima prediksi berdasarkan data klinis, meningkatkan deteksi dini dan kesadaran akan resiko hepatitis.

**Abstract.** Hepatitis is a contagious disease that has become a leading cause of death worldwide. This research proposes the use of the CART decision tree method to detect hepatitis disease early. The model is trained and evaluated using clinical data to identify patterns and significant indicators in hepatitis diagnosis. The evaluation results show a high accuracy of 99.17%, indicating the model's effectiveness in detecting hepatitis. The implementation of the model into a website allows users to receive predictions based on clinical data input, enhancing early detection and awareness of hepatitis risks.

## 1. PENDAHULUAN

Salah satu negara dengan jumlah penderita penyakit yang sangat tinggi ialah Indonesia. Di antara berbagai penyakit yang umum terjadi, Penyakit yang paling banyak diderita oleh masyarakat Indonesia salah satunya ialah Hepatitis [1]. Sekitar 2 miliar orang di seluruh dunia terinfeksi hepatitis berdasarkan data dari World Health Organization (WHO), dan 1,4 juta di antaranya meninggal setiap tahunnya akibat penyakit ini.[2]. Perkembangan teknologi saat ini sudah sangat pesat dan juga

memberikan kontribusi yang besar terhadap pengembangan sistem yang dapat melakukan pendeteksian dini penyakit hepatitis. Sistem ini memiliki informasi yang dapat memproses data dan informasi yang didapatkan untuk melakukan pendeteksian dini penyakit hepatitis menjadi lebih cepat dan lebih akurat.

Untuk sistem pendeteksi dini penyakit hepatitis dalam penelitian ini kami mengusulkan penggunaan metode CART decision tree untuk mendeteksi penyakit hepatitis. Pendekatan ini diharapkan dapat

menyediakan alat diagnostik yang cepat, akurat, dan mudah diimplementasikan, yang dapat membantu tenaga medis dalam membuat keputusan klinis yang lebih baik. Dengan menganalisis data medis pasien, seperti hasil tes darah dan gejala klinis, model CART decision tree dapat mengidentifikasi pola dan indikator yang signifikan dalam diagnosis hepatitis[3].

Penelitian ini akan mencakup beberapa tahap, mulai dari pengumpulan dan pra-pemrosesan data, pembangunan dan pelatihan model CART decision tree, hingga evaluasi kinerja model dengan menggunakan metrik seperti akurasi, sensitivitas, dan spesifisitas. Hasil dari penelitian ini diharapkan tidak hanya memperkuat literatur ilmiah mengenai aplikasi pembelajaran mesin dalam deteksi penyakit, tetapi juga memberikan solusi praktis untuk peningkatan kualitas layanan kesehatan dalam diagnosis hepatitis.

Dengan demikian, penelitian ini bertujuan untuk mengembangkan sistem yang dapat mendiagnosis penyakit hepatitis menggunakan CART decision tree dan untuk mengetahui seberapa efektif sistem tersebut dalam mendiagnosis penyakit hepatitis. Diharapkan hasil penelitian ini dapat membantu meningkatkan *awareness* tentang risiko penyakit hepatitis dan cara mencegahnya, serta membantu meningkatkan akses ke fasilitas medis yang memadai untuk mendiagnosis dan mengobati penyakit hepatitis.

## **2. TINJAUAN PUSTAKA**

### **2.1. Sistem Pendukung Keputusan**

Sistem Pendukung Keputusan (SPK) adalah suatu sistem yang dirancang guna membantu pengambilan keputusan dalam kondisi yang semiterstruktur dan kondisi yang tidak terstruktur [4]. Sistem ini menggunakan data dan model analisis canggih untuk memproses informasi dan memberikan rekomendasi yang lebih tepat.

SPK dapat membantu pengambilan keputusan dengan cara menghimpun data yang relevan, menganalisa data, serta memberikan rekomendasi dimana didasarkan pada analisis tersebut [5]. Sistem ini dapat digunakan dalam berbagai bidang, seperti bisnis, manajemen, dan

keuangan, untuk membantu mengambil keputusan yang lebih efisien serta efektif.

### **2.2. Hepatitis**

Hepatitis adalah kondisi yang menyebabkan peradangan di hati ataupun liver. Berbagai faktor yang dapat menyebabkan kondisi ini ialah seperti penggunaan obat-obatan tertentu, konsumsi alkohol berlebihan, infeksi virus, infeksi cacing hati, serta penyakit autoimun. Penyakit ini dapat menular ke orang lain ketika infeksi virus menjadi penyebab hepatitis [6]. Penyakit hepatitis dapat menular ke orang lain dan dapat berlangsung selama beberapa bulan atau lebih dari enam bulan, tergantung pada jenis dan tingkat keparahan penyakit.

Terdapat lima jenis penyakit hepatitis, yaitu Hepatitis A, Hepatitis B, Hepatitis C, Hepatitis D, serta Hepatitis E[2].

#### **2.2.1. Hepatitis A**

Hepatitis A merupakan jenis hepatitis dimana disebabkan oleh infeksi virus hepatitis A (HAV). Makanan atau minuman yang terkontaminasi virus hepatitis A menjadi penyebab penularan penyakit ini [6].

#### **2.2.2. Hepatitis B**

Hepatitis B adalah penyakit dimana disebabkan oleh infeksi virus hepatitis B (HBV). Penyakit ini bisa menyebar melalui pemindahan darah serta kontak seksual tanpa pelindung [6].

#### **2.2.3. Hepatitis C**

Hepatitis C disebabkan oleh infeksi virus hepatitis C (HCV). Penularan penyakit ini bisa terjadi dengan melakukan melalui pemakaian jarum suntik yang tidak steril atau hubungan seksual tanpa penggunaan kondom [6].

#### **2.2.4. Hepatitis D**

Hepatitis D merupakan penyakit yang menyebabkan peradangan pada hati efek dari infeksi virus hepatitis D (HDV). Meskipun jarang terjadi, hepatitis D dapat menyebabkan masalah kesehatan yang serius. Penularan penyakit ini dapat terjadi melalui transfusi darah ataupun penggunaan jarum suntik yang tidak steril [6].

### 2.2.5. Hepatitis E

Hepatitis E merupakan penyakit dimana disebabkan oleh infeksi virus hepatitis E (HEV). Konsumsi air ataupun makanan yang terkontaminasi virus tersebut menjadi penyebab penularan penyakit ini [6].

### 2.3. Algoritma

Algoritma merupakan serangkaian tahapan dimana dirangkai dengan logis serta sistematis guna mengatasi suatu permasalahan. Algoritma ialah tahapan pertama yang ditulis sebelum pembuatan program dimulai dalam pemrograman komputer dasar. Salah satu jenis masalah yang bisa diatasi menggunakan pemrograman komputer ialah masalah yang menyangkut penghitungan matematis [7].

Algoritma dapat diterapkan di berbagai bidang, seperti pemrograman, ilmu komputer, serta dalam kehidupan sehari-hari demi menyelesaikan permasalahan yang memerlukan prosedur atau proses tertentu[8].

### 2.4. Cart Decision Tree

Classification and Regression Trees (CART) adalah sebuah algoritma ataupun metode dalam teknik pencarian data yang memanfaatkan pendekatan pohon keputusan. Metode sangat efektif meskipun dikenal sederhana. CART bertujuan untuk mengidentifikasi klasifikasi data secara seksama dalam klasifikasi tertentu. Selain itu, CART digunakan untuk memodelkan korelasi antara variabel dependen (variabel respons) dan satu ataupun lebih variabel independen (variabel prediktor)[9].

Pohon klasifikasi akan dihasilkan oleh CART asalkan variabel respons berskala kategorikal, serta jika variabel respons berupa data kontinu akan menghasilkan pohon regresi. CART bertujuan untuk membagi data secara tepat menjadi kelompok-kelompok yang dapat berfungsi sebagai pengenalan yang andal untuk proses klasifikasi[9].

## 3. METODE PENELITIAN

### 3.1. Studi Literatur

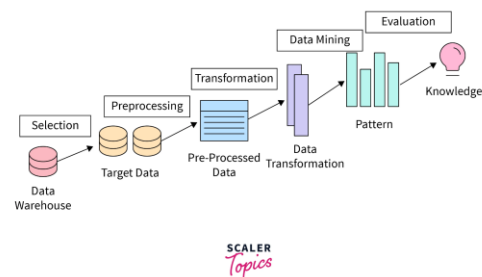
Penyelesaian masalah melalui penelusuran berbagai sumber tulisan yang telah ada sebelumnya merupakan cara melakukan studi literatur. Dalam konteks ini, pengalaman

subjek menjadi fenomena utama yang menjadi fokus penelitian[10]. Proses ini melibatkan pencarian dan identifikasi literatur yang mencakup artikel ilmiah, laporan penelitian, dan sumber-sumber lainnya yang terkait. Setelah data dan informasi tersebut dikumpulkan, tahapan berikutnya ialah menganalisa berbagai data dimana diperoleh untuk memahami konteks, menemukan pola, dan mengidentifikasi kesenjangan dalam pengetahuan yang ada.

### 3.2. Knowledge Discovery in Database

Proses KDD (Knowledge Discovery in Database) merupakan metode yang digunakan demi menjalankan data mining, di mana KDD melibatkan serangkaian langkah untuk mengeksplorasi, menganalisis, serta mengekstrak data dengan demikian dapat diubah menjadi sumber yang bermanfaat [11].

Penelitian ini menggunakan metode KDD (Knowledge Discovery in Database). KDD memiliki lima tahapan utama, dimana meliputi Data Selection, Preprocessing, Transformation, Data Mining, serta Evaluation. Setiap tahapan ini memainkan peran penting dalam proses ekstraksi pengetahuan dari data [12].



Gambar 1. Tahap KDD

#### 3.2.1. Data Selection

Data yang relevan dengan penelitian dipilih dari keseluruhan dataset pada langkah Data Selection. Dalam konteks ini, data yang dipilih mencakup fitur-fitur gejala seperti 'itching', 'skin\_rash', dan lainnya, serta hasil diagnosis ('prognosis'). Pemilihan ini memastikan bahwa hanya data yang berkaitan langsung dengan prediksi penyakit yang

digunakan untuk analisis selanjutnya, dengan menghapus kolom yang tidak diperlukan dan memastikan data tersebut bersih dari nilai yang hilang.

### 3.2.2. Pre-processing

Pada tahap Preprocessing, data yang telah dipilih melalui tahap Data Selection kemudian dibersihkan dan dipersiapkan untuk analisis lebih lanjut. Proses ini meliputi penanganan nilai yang hilang, normalisasi data, serta transformasi variabel agar sesuai dengan kebutuhan algoritma yang akan digunakan. Tujuan dari preprocessing adalah untuk memastikan bahwa data dalam kondisi optimal dan bebas dari noise, sehingga hasil analisis menjadi lebih akurat dan andal.

### 3.2.3. Transformation

Pada tahap Transformation, data yang telah melalui preprocessing diubah atau dimodifikasi agar lebih sesuai dengan kebutuhan analisis. transformasi dilakukan dengan mengubah struktur data atau formatnya agar dapat digunakan dengan algoritma machine learning decision tree.

### 3.2.4. Data Mining

Pada tahap Data Mining, algoritma digunakan untuk menemukan pola atau membuat prediksi berdasarkan data yang telah diproses. Data Mining dilakukan dengan menggunakan algoritma Decision Tree Classifier.

Algoritma ini dilatih dengan data yang telah dipilih dan dipreproses untuk memprediksi hasil diagnosa penyakit berdasarkan gejala yang ada. Model decision tree ini kemudian menghasilkan pohon keputusan yang digunakan untuk klasifikasi data baru, memungkinkan identifikasi pola dalam dataset yang dapat digunakan untuk prediksi lebih lanjut.

### 3.2.5. Evaluation

Pada tahap Evaluation, model yang dihasilkan dari Data Mining dievaluasi untuk menilai kinerjanya. Evaluasi dilakukan dengan menghitung metrik seperti *accuracy score* dan membuat *classification report*. Metrik-metrik ini digunakan untuk menilai seberapa baik model decision tree dalam memprediksi hasil diagnosa penyakit.

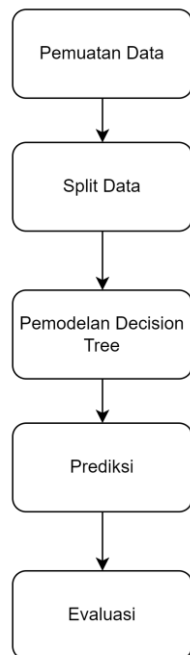
## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Dataset

Penelitian ini memiliki dataset dimana terdiri dari atribut yang mencakup karakteristik klinis terkait dengan penyakit hepatitis. Terdapat dua belas fitur yang ada di dalam dataset yang merupakan gejala - gejala dari setiap jenis penyakit hepatitis.

Selain itu, dataset juga mencakup label kelas (target) yaitu "prognosis", yang menyediakan informasi tentang diagnosis atau prognosis penyakit hepatitis. Jumlah data yang dipergunakan dalam penelitian sejumlah 600 yang datanya diambil dari Kaggle.com. Dari total data tersebut, 480 data digunakan untuk pelatihan dan 120 data digunakan untuk pengujian model. Kasus klasifikasi yang ditangani adalah penyakit Hepatitis A, Hepatitis B, Hepatitis C, Hepatitis D, dan Hepatitis E.

## 4.2. Proses Penerapan Algoritma



Gambar 2. Alur Proses Penerapan Algoritma

Berikut merupakan penjelasan dari setiap proses penerapan algoritma yang dilakukan:

### 4.2.1. Pemuatan Data

Data yang relevan untuk penelitian dimuat ke dalam lingkungan komputasi pada proses ini. Ini termasuk proses membaca data dari sumber aslinya, seperti file CSV atau database, dan memuatnya ke dalam struktur data yang sesuai, seperti DataFrame dalam lingkungan pemrograman Python menggunakan library pandas.

### 4.2.2. Split Data

Setelah data dimuat, langkah selanjutnya adalah membaginya menjadi dua subset yang saling eksklusif: set pelatihan dan set pengujian. Set pelatihan diterapkan guna melatih model, sementara set pengujian diterapkan guna melakukan uji pada performa model yang sudah dilatih. Klasifikasi ini penting untuk menghindari bias dan memastikan generalisasi model terbaik pada data yang belum pernah ditelaah sebelumnya.

### 4.2.3. Pemodelan Decision Tree

Proses ini melibatkan konstruksi model Decision Tree menggunakan data pelatihan. Decision Tree merupakan metode pembelajaran mesin dimana memetakan rentetan keputusan

berjenjang berdasarkan fitur-fitur dalam data ke label kelas yang sesuai. Model ini dikonstruksi dengan mengidentifikasi aturan-aturan yang memisahkan kelas-kelas yang berbeda berdasarkan nilai-nilai fitur yang ada dalam dataset.

### 4.2.4. Prediksi

Setelah model Decision Tree dibangun, langkah selanjutnya adalah menggunakan model tersebut untuk memprediksi label kelas pada data pengujian yang sebelumnya tidak terlihat. Proses ini melibatkan memberikan input dari setiap sampel data pengujian ke model, yang kemudian menghasilkan prediksi kelas yang sesuai berdasarkan aturan yang telah ditemukan selama pelatihan.

### 4.2.5. Evaluasi

Evaluasi model dilakukan dengan membandingkan prediksi yang dihasilkan oleh model dengan label kelas yang sebenarnya pada data pengujian. Menilai performa model dapat menggunakan berbagai matrik penilaian seperti F1-Score, akurasi, recall, serta presisi. Evaluasi ini menyuguhkan pemahaman mengenai seberapa baik model dapat digunakan untuk memprediksi label kelas pada data baru dan relevansi serta keandalannya dalam aplikasi praktis.

```

In [15]: from sklearn.metrics import classification_report

# Menghitung presisi, recall, dan F1-Score
report = classification_report(y_test, pred)
print(report)

```

	precision	recall	f1-score	support
Hepatitis B	1.00	1.00	1.00	25
Hepatitis C	1.00	1.00	1.00	18
Hepatitis D	1.00	0.96	0.98	27
Hepatitis E	1.00	1.00	1.00	27
hepatitis A	0.96	1.00	0.98	23
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

Gambar 3. Hasil Evaluasi Pemodelan

Dalam evaluasi model Decision Tree yang telah dilatih, hasil berdasarkan Gambar 3 menunjukkan performa yang luar biasa dalam mengklasifikasikan jenis-jenis Hepatitis pada dataset pengujian. Dengan tingkat akurasi sebesar 99%, model ini mampu memberikan prediksi yang benar untuk hampir semua kasus yang diuji. Presisi tinggi sebesar 99% menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan oleh model adalah benar, sementara recall sebesar 99% menunjukkan kemampuan model dalam

mengidentifikasi hampir semua instance positif yang sebenarnya ada dalam data pengujian. F1-Score yang juga mencapai 99% menegaskan bahwa model ini berhasil mempertahankan keseimbangan yang baik antara recall dan presisi, memperkuat kualitas seluruh model dalam memperhitungkan label kelas.

Dengan hasil evaluasi ini, dapat disimpulkan bahwa model Decision Tree yang dikembangkan memiliki kemampuan yang sangat baik dalam mendeteksi berbagai jenis Hepatitis berdasarkan data gejala yang tersedia. Hasil ini menunjukkan potensi penggunaan model ini dalam aplikasi klinis untuk diagnosis penyakit secara lebih akurat.

#### 4.3. Implementasi menjadi Website

##### 4.3.1. Halaman Form

Halaman Form ditampilkan ketika pengguna memasuki web pertama kali. Pada halaman Form berisikan mengenai pertanyaan - pertanyaan yang pertanyaan tersebut menjadi penilaian model untuk menghasilkan keputusan.

##### Aplikasi Prediksi Hepatitis

Masukkan Gejala Anda:

Nota: 0 = Tidak, 1 = Ya

Apakah Anda Merasakan Gatal-gatal?	Apakah Anda Merasakan Henti Sendi?	Apakah Anda Merasakan Mual?	Apakah Anda Merasakan Lelah?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Apakah Anda Merasakan Lethargi?	Apakah Anda Merasakan Demam Tinggi?	Apakah Anda Merasakan Kulit Kuning?	Apakah Anda Merasakan Urine Gelap?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Apakah Anda Merasakan Hilangnya Nafsu Makan?	Apakah Anda Merasakan Henti Perut?	Apakah Anda Merasakan Urine Kuning?	Apakah Anda Merasakan Mata Kuning?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Apakah Anda Merasakan Gagal Ginjal?	Apakah Anda Merasakan Malaise?	Apakah Ada Bileage Keluar?	Apakah Anda Pernah Transfusi Darah?
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Apakah Anda Pernah Suntik Steril?	Apakah Anda Pernah Merasakan Koma?	Apakah Anda Pernah Pendarahan Lambung?	
<input type="text"/>	<input type="text"/>	<input type="text"/>	

Prediksi

Gambar 4. Halaman Form Website.

##### 4.3.2. Tampilan Hasil

Setelah pengguna melakukan pengisian pertanyaan pada form yang diberikan, maka akan muncul *output* yang dikeluarkan di bawah tombol prediksi Hasil prediksi berupa text yang menunjukan penyakit hepatitis yang kemungkinan dideritanya.

Prediksi

Anda Diprediksi Terkena Hepatitis A

Gambar 5. Tampilan Hasil Deteksi

#### 5. KESIMPULAN

- Model Decision Tree CART yang dikembangkan dalam penelitian ini menunjukkan performa yang sangat baik dalam memprediksi label kelas pada dataset pengujian penyakit hepatitis.
- Implementasi model menjadi website dengan halaman dashboard dan tampilan hasil memungkinkan pengguna untuk memperoleh prediksi penyakit hepatitis berdasarkan pertanyaan yang dijawab.

#### DAFTAR PUSTAKA

- [1] Syahputra H, Syafindy DM. Sistem Pakar Diagnosa Penyakit Hepatitis Dengan Menggunakan Metode Certainty Factor. Jurnal Sains Informatika Terapan. 2023;2(1):45-50.
- [2] V. I. P., "Hepatitis: Jenis, Penyebab, Gejala, Dan Pengobatan," Direktorat Jenderal Pelayanan Kesehatan, [https://yankes.kemkes.go.id/view\\_artikel/1993/hepatitis-jenis-penyakit-gejala-dan-pengobatan](https://yankes.kemkes.go.id/view_artikel/1993/hepatitis-jenis-penyakit-gejala-dan-pengobatan) (accessed Aug. 5, 2024).
- [3] Jones AH, Makmun MS. Implementasi Metode CART untuk Klasifikasi Diagnosis Penyakit Hepatitis Pada Anak. Journal of Informatics Information System Software Engineering and Applications (INISTA). 2021 Jul 9;3(2):61-70.
- [4] Rifqi A, Aldisa RT. Implementasi Sistem Pendukung Keputusan dalam Pemilihan Vendor IT Menerapkan Metode Weighted Product (WP). Resolusi: Rekayasa Teknik Informatika dan Informasi. 2023 Sep 30;4(1):1-7.
- [5] M. SMK, "Memahami Sistem Pendukung Keputusan (DSS): Pengertian, Komponen, Dan Manfaatnya," Diploma of Telecommunication Engineering, <https://dte.telkomuniversity.ac.id/memahami-sistem-pendukung-keputusan-dss-pengertian-komponen-dan-manfaatnya/> (accessed Aug. 6, 2024).
- [6] Pittara, "Hepatitis," Alodokter, <https://www.alodokter.com/hepatitis> (accessed Aug. 7, 2024).
- [7] Rangkuti A, Yahfizham Y. Pengenalan Algoritma Pemrograman Dasar Dalam Konteks Pembelajaran Pemrograman Awal. Konstanta: Jurnal Matematika Dan Ilmu Pengetahuan Alam. 2023 Nov 10;1(4):223-37.
- [8] Nasution ZM, Siadari MM, Saragih IJ, Kirana IO, Siregar ZA. Penerapan Matematika Algoritma dalam Bidang Komputer. FARABI: Jurnal Matematika dan Pendidikan Matematika. 2023 Dec 20;6(2):180-91.

- [9] Siregar AS, Siregar YS, Khairani M. Implementation Of The Data Mining Cart Algorithm In The Characteristic Pattern Of New Student Admissions. *Journal of Computer Networks, Architecture and High Performance Computing*. 2023 Feb 21;5(1):263-75.
- [10] Nuryana A, Pawito P, Utari P. Pengantar metode penelitian kepada suatu pengertian yang mendalam mengenai konsep fenomenologi. *Ensains Journal*, 2 (1), 19–24.
- [11] Buani DC. Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest. *EVOLUSI: Jurnal Sains dan Manajemen*. 2024 Jun 13;12(1).
- [12] Nurian A. Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*. 2023 Sep 12;11(3s1).