

ALGORITMA C4.5 UNTUK MEMPREDIKSI JUMLAH PENGGUNAAN BAHAN BAKAR TERHADAP JUMLAH KENDARAAN TEMPUR MILITER PADA SUATU NEGARA

Ngawal Muhamad, Rizal Adi Saputra

Teknik Informatika, Universitas Halu Oleo

Kampus Bumi Hijau Tridharma Anduonohu Jalan H.E.A. Mokodompit

rizaladisaputra@uho.ac.id

ABSTRAK

Penggunaan bahan bakar kendaraan menjadi salah satu pengeluaran terbesar di dunia. Terkhusus pada lingkup militer penggunaan bahan bakar menjadi objek vital dan menjadi sumber daya utama dalam peralatan dan pada saat pelaksanaan operasi-operasi militer tertentu. Oleh karena itu penelitian ini dilakukan dengan tujuan agar dapat mengetahui jumlah penggunaan bahan bakar militer pada suatu negara berdasarkan jumlah kendaraan militernya. Sehingga membantu negara untuk mengoptimalkan sumber daya bahan bakarnya dan merencanakan strategi baru yang lebih efektif. Metode yang digunakan pada penelitian ini adalah metode algoritma C4.5 yang merupakan salah satu algoritma untuk menentukan suatu klasifikasi kelompok data dan bersifat prediktif. Adapun kelebihan dari algoritma ini adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima. Hasil dari penelitian ini adalah dapat memprediksi jumlah penggunaan bahan bakar terhadap jumlah kendaraan militer dari suatu negara di dunia. Data yang digunakan dalam penelitian ini berasal dari situs web *globalfirepower.com* dengan sebanyak 144 data negara. Hasil dari penelitian ini menggunakan bantuan tools berupa situs web *google colab* dengan mengimpor model-model dan pustaka dari python dan menghasilkan tingkat akurasi sebesar 96%.

Kata kunci : algoritma C4.5, jumlah kendaraan tempur, prediksi, penggunaan bahan bakar

1. PENDAHULUAN

Dalam konteks militer, istilah "militer" mengacu pada segala sesuatu yang berhubungan dengan angkatan bersenjata atau kekuatan militer suatu negara. Ini mencakup organisasi, personel, peralatan, strategi, dan operasi yang berkaitan dengan keamanan dan pertahanan nasional. Fungsi militer antara lain adalah melindungi negara dari ancaman eksternal dan menjaga keamanan dalam negeri. Salah satu aspek penting dalam operasi militer adalah penggunaan bahan bakar kendaraan tempur. Efisiensi penggunaan bahan bakar ini sangat penting untuk logistik dan perencanaan operasional suatu negara.

Penggunaan bahan bakar untuk kendaraan tempur militer merupakan salah satu komponen utama dalam anggaran militer. Oleh karena itu, pengelolaan penggunaan bahan bakar yang efisien dapat memberikan kontribusi besar dalam penghematan biaya dan peningkatan efektivitas operasional. Untuk mencapai hal ini, diperlukan metode analisis data yang canggih dan akurat. Salah satu metode yang dapat digunakan adalah algoritma C4.5.

Algoritma C4.5 adalah algoritma pembelajaran mesin yang digunakan untuk membangun pohon keputusan dari data pelatihan. Pohon keputusan ini kemudian digunakan untuk klasifikasi dan prediksi. Algoritma ini terbukti efektif dalam analisis data dan mampu menghasilkan pohon keputusan yang memiliki tingkat akurasi yang tinggi.

Decision tree atau pohon keputusan adalah salah satu metode dalam data mining yang sangat berguna untuk analisis data dan pengambilan keputusan. Pohon keputusan memungkinkan pemodelan keputusan dan konsekuensi yang mungkin, termasuk peluang

kejadian, biaya, dan utilitas. Dalam *data mining*, *decision tree* digunakan untuk menemukan pola-pola penting dalam data dan membantu dalam klasifikasi serta prediksi.

Permasalahan utama yang dibahas dalam penelitian ini adalah bagaimana memprediksi jumlah penggunaan bahan bakar kendaraan tempur militer pada suatu negara berdasarkan jumlah kendaraan tempur yang dimiliki. Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang akurat mengenai konsumsi bahan bakar kendaraan tempur militer menggunakan algoritma C4.5. Model ini diharapkan dapat membantu dalam pengambilan keputusan terkait alokasi bahan bakar dan perencanaan strategi operasional militer. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam bidang ilmu pengetahuan data dan pembelajaran mesin, tetapi juga memberikan manfaat praktis bagi pengelolaan sumber daya militer suatu negara.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining adalah kandungan ikhtisar dari kelompok data yang bertujuan untuk menggali informasi yang bermanfaat di dalam data tersebut, dimana informasi berharga tersebut tidak diketahui sebelumnya [1].

Data Mining bukanlah sekumpulan data biasa yang mencakup analisis dan prediksi yang ditampilkan saja, namun data tersebut dikumpulkan dan disimpan ke dalam *database* yang diproses untuk dijadikan sample pengambilan keputusan. *Data mining* merupakan suatu proses penggalian atau penambangan informasi untuk menemukan informasi yang tidak

ditemukan sebelumnya dari sekumpulan data besar (big data) atau repository database lainnya [2].

Data Mining melakukan Proses penyingkapan atau ekstraksi data menjadi pengetahuan dimulai dari data yang tidak memiliki arti, kemudian diolah menjadi informasi berupa rangkuman dan statistik data [3].

Dalam jurnal ilmiah, *data mining* juga dikenal dengan nama KDD (*Knowledge Discovery in Database*). Namun pada tahun 1995, telah diadakan *International KDD Conference* di Montreal yang mendefinisikan bahwa KDD adalah proses untuk mengenali informasi atau suatu kebenaran baru dan benar-benar berguna dengan mengenali pola dari data. Tujuan utama dari proses KDD adalah memprediksi nilai-nilai yang berguna dari variabel-variabel dengan menemukan pola-pola dari sebuah gagasan data yang diinterpretasikan oleh manusia [4].

2.2. Decision Tree

Decision Tree merupakan diagram berbentuk struktur pohon yang setiap internal node menyatakan pengujian terhadap suatu atribut, dan setiap cabang menyatakan output dari pengujian serta leaf node (bagian daun) menyatakan kelas-kelas atau distribusi kelas. Node yang berada paling atas dari *Decision Tree* disebut sebagai *rootnode* atau *node akar*, yang melahirkan leaf node atau hasil akhir yang mewakili label kelas dari kombinasi atribut sehingga terbentuk menjadi *rule* (aturan) [1].

Pohon keputusan merupakan metode klasifikasi dan prediksi yang kuat dan terkenal. Pada metode pohon keputusan tersebut data diubah menjadi fakta besar menjadi pohon keputusan yang mempresentasikan aturan-aturan. Aturan tersebut dapat dengan mudah dipahami ataupun dijelaskan dengan ulang karena memiliki bahasa yang dapat dimengerti. Pohon keputusan juga berguna untuk menjelajahi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. "Pohon keputusan memiliki fleksibilitas (mudah digunakan yang membuatnya sangat menarik, terutama karena menyajikan keuntungan dari proses perubahan konsep menjadi gambar yang telah disajikan sangat mudah dimengerti dimana cabang-cabang dari pohon merangkum klasifikasi" [5].

Decision tree merupakan metode klasifikasi yang pada prinsipnya adalah membuat struktur pohon seperti diagram alur di mana setiap *node (node non leaf)* mewakili atribut dan setiap cabang pada pohon yang mewakili hasil uji, serta *node leaf* menunjukkan kelas[6].

Decision tree juga merupakan teknik klasifikasi dan prediksi yang mampu mengubah masalah besar menjadi pohon keputusan dengan aturan [7].

2.3. Algoritma C4.5

Algoritma C4.5 merupakan metode yang menjadi pilihan utama untuk pengembangan data mining karena kecepatannya dalam klasifikasi data atau

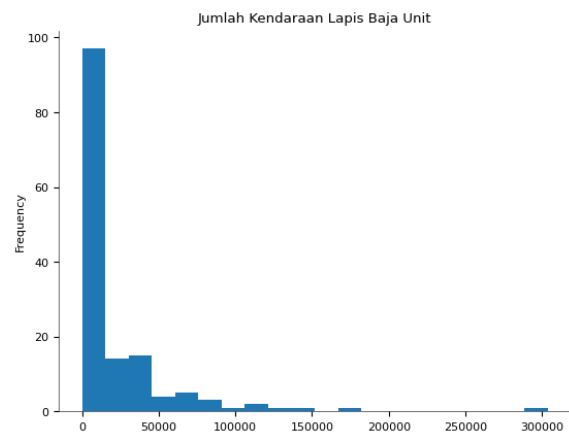
segmentasi dan bersifat prediktif. Algoritma ini merupakan salah satu metode yang menghasilkan *Decision Tree* berdasarkan training data. Kelebihannya adalah metode pohon keputusan yang dapat menghindari munculnya permasalahan dengan kriteria yang jumlahnya lebih sedikit pada node internal tanpa menguraingi keputusan yang dihasilkan [5].

Algoritma ini berfokus pada klasifikasi data bertipe pohon keputusan. Pohon keputusan Algoritma C4.5 dibangun dengan beberapa tahap yang meliputi pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan membagi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [8].

Algoritma C4.5 memiliki inputan berupa *training samples* dimana data yang digunakan untuk membangun sebuah pohon keputusan yang telah diuji kebenarannya dan *samples* adalah *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan pengklasifikasian data [9].

Algoritma C4.5 ialah salah satu metode klasifikasi yang merepresentasikan hasil dalam struktur pohon (*tree*) yang mana setiap node merepresentasikan nilai atribut, cabangnya merepresentasikan nilai dari atribut serta daun kelas [10].

2.4. Gambar Grafik



Gambar 1. Jumlah kendaraan lapis baja

Dari grafik pada gambar 1, dapat dilihat :

- Terdapat dua kolom dalam tabel tersebut, yaitu "*Frequency*" dan "*Jumlah Kendaraan Lapis Baja Unit*".
- Frekuensi kendaraan dengan lapisan baja sebanyak 100 unit terdapat pada kisaran jumlah kendaraan sebanyak 50,000 hingga 200,000 unit.
- Frekuensi kendaraan dengan lapisan baja sebanyak 20 unit terdapat pada jumlah kendaraan sebanyak 250,000 hingga 300,000 unit.

2.5. Tabel Dataset

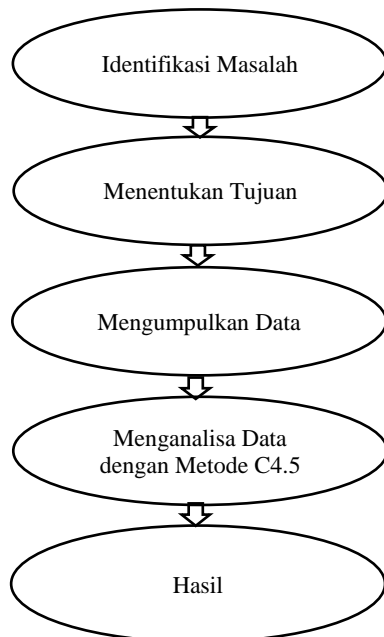
Tabel 1. Tabel Dataset

No	Negara	Jumlah Kendaraan Lapis Baja	Konsumsi Bahan Bakar	Kelas
1	Amerika	303553	20555000	Tinggi
2	Rusia	151641	3700000	Sedang
3	China	174300	14010000	Tinggi
4	India	100882	5000000	Sedang
...
140	Liberia	650	9500	Rendah
141	Somalia	3712	6000	Rendah
142	Moldova	550	22000	Rendah
143	Benin	150	45000	Rendah
144	Bhutan	108	4500	Rendah

Tabel tersebut menyajikan data terkait jumlah kendaraan lapis baja, konsumsi bahan bakar, dan kelas penggunaan bahan bakar untuk kendaraan tempur militer di beberapa negara. Setiap baris mewakili satu negara, dengan kolom "No" menunjukkan urutan, "Negara" menyebutkan nama negara tersebut, "Jumlah Kendaraan Lapis Baja" mencantumkan total kendaraan tempur militer yang dimiliki, "Konsumsi Bahan Bakar" mencatat jumlah bahan bakar yang dikonsumsi, dan "Kelas" menunjukkan kategori penggunaan bahan bakar (Tinggi, Sedang, atau Rendah).

3. METODE PENELITIAN

Metodologi Penelitian merupakan tahap-tahap penelitian yang sistematis untuk membantu penelitian menjadi terarah dengan baik. Berikut adalah metodologi penelitian yang dilakukan.



Gambar 2. Tahapan penelitian

Tahapan penelitian seperti pada Gambar 2 memiliki 5 tahapan, yaitu :

- a. Identifikasi Masalah
Merupakan tahapan awal dari sebuah penelitian, dimana peneliti akan membuat rumusan masalah lalu menentukan batasan dari permasalahan yang diteliti agar lebih terarah.
- b. Menentukan Tujuan
Tahapan ini dilakukan agar peneliti tidak menyimpang dari tujuan yang ingin dicapai
- c. Mengumpulkan Data
Pengumpulan data yang dilakukan oleh peneliti dengan melakukan observasi dan mencari data pada situs web globalfirepower.com
- d. Menganalisa Data Dengan Metode Algoritma C4.5
Agar permasalahan pada penelitian ini dapat di analisa maka digunakan metode algoritma C4.5 dengan tahanannya adalah mempersiapkan dataset berbentuk format csv, kemudian mengimportkan data pada *google colab* dengan menggunakan *tools* panda lalu melakukan penyandingan pada kolom dataset dan membuat model pohon keputusan algoritma C4.5
- e. Hasil
Hasil dari penelitian adalah berupa pohon keputusan yang disimpan dalam bentuk grafis.

4. HASIL DAN PEMBAHASAN

Dalam penelitian, data sampel yang digunakan adalah 144 data negara *power index* dari situs globalfirepower.com. Setelah data yang diperlukan diperoleh, kemudian ditentukan atribut-atribut yang diperlukan pada penelitian ini antara lain : Nama Negara, Jumlah Kendaraan Lapis Baja (Unit), Jumlah Penggunaan Bahan Bakar (bbl) dan Kelas yang mewakili label dalam data lalu dimasukan ke dalam dokumen excel sebagai dataset dan di *import* pada *Google Colab* untuk diolah dengan menggunakan perhitungan otomatis algoritma C4.5 .

Data yang telah di import selanjutnya akan diberi penyandian atau *encoding* pada kategori atau label menjadi bilangan bulat pada kolom 'Negara' dan 'Kelas' di dalam dataset menggunakan objek *encoder* untuk merubah nilai-nilai kategorikal menjadi nilai numerik

Data yang telah di import selanjutnya akan diberi penyandian atau *encoding* pada kategori atau label menjadi bilangan bulat pada kolom 'Negara' dan 'Kelas' di dalam dataset menggunakan objek *encoder* untuk merubah nilai-nilai kategorikal menjadi nilai numerik

Setelah penyandian atau encoding selesai dataset akan menghapus kolom 'Kelas' dan menyimpannya ke dalam variabel dengan perintah `atr_dataset = dataset.drop(columns = 'Kelas')`. Kolom 'Kelas' kemudian diambil dari *DataFrame* dataset dan menyimpannya ke dalam variabel `cls_dataset`.

Modul-modul pada pustaka *mechine learning* di python diimport, untuk melatih model *Decission Tree* pada data pelatihan. Prediksi kemudian dilakukan menggunakan model pohon keputusan C4.5 dan menampilkannya menjadi *Confussion Matrix* dan menghasilkan file DOT yang merepresentasikan

struktur pohon keputusan dari model yang telah dilatih menggunakan dataset (tree_dataset)

Tabel 3. Tampilan Confussion Matrix

Confusion Matrix					
[[26 0] [1 2]]					
Tingkat Akurasi Algoritma C4.5					
Akurasi		Precision	recall	f1-score	support
0	0.96	1.00	0.98		26
2	1.00	0.96	0.80		3
accuracy			0.97		29
Macro avg	0.98	0.83	0.89		29
Weighted avg	0.97	0.97	0.96		29
Tingkat akurasi : 96 persen					

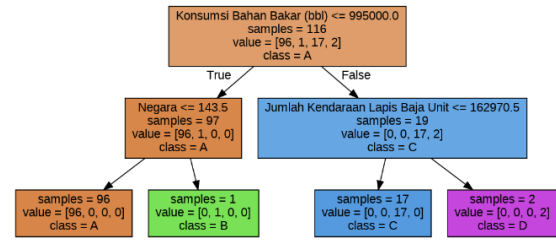
Tabel tersebut menampilkan representasi dari pohon keputusan dengan bentuk tampilan *confussion matrix*. Dengan hasil tingkat akurasi sebesar 96 persen.

Tabel 4. Tabel hasil dot

```

Diagraph Tree {
node [shape = box, style="filled", color="black",
fontname"helvetiva"];
edge [fontname = "helvetica"];
0 [label="konsumsi Bahan Bakar (bb) <=
995000.0\nsamples = 116\nvalue = [96, 1, 17, 2]\nclass =
A", fillcolor="#ea9a61"];
1 [label="Negara <= 143.5\nsamples = 97\nvalue =
[96,1,0,0]\nclass = A", fillcolor = #e5823b"];
0 -> 1 [labeldistance=2.5, labelangle=45,
headlabel="True"];
2 [label="samples = 96\nvalue = [96, 0, 0, 0]\nclass =
A", fillcolor="#e58139"];
1 -> 2;
3 [label="samples = 1\nvalue = [0, 1, 0, 0]\nclass = B",
fillcolor="#47e539"];
1 -> 3;
4 [label="Jumlah Kendaraan Lapis Baja Unit <=
162970.5\nsamples = 19\nvalue = [0, 0, 17, 2]\nclass =
C", fillcolor="#50a9e8"];
0 -> 4 [labeldistance=2.5, labelangle=-45,
headlabel="False"];
5 [label="samples = 17\nvalue = [0, 0, 17, 0]\nclass = C",
fillcolor="#399de5"];
4 -> 5;
6 [label="samples = 2\nvalue = [0, 0, 0, 2]\nclass = D",
fillcolor="#d739e5"];
4 -> 6;
}
    
```

Tabel tersebut adalah tampilan hasil file DOT dari representasi pohon keputusan dari model yang dilatih menggunakan dataset



Gambar 3. Hasil pohon keputusan dalam bentuk grafis

Analisis dari diagram pohon keputusan tersebut dalam menggunakan algoritma C4.5 untuk memprediksi jumlah penggunaan bahan bakar terhadap jumlah kendaraan tempur militer pada suatu negara adalah sebagai berikut:

- Node Root**
Pohon keputusan dimulai dengan atribut "Konsumsi Bahan Bakar (bb)" dengan batas 995000.0. Dari 116 sampel, 96 sampel masuk kelas A, 1 sampel masuk kelas B, 17 sampel masuk kelas C, dan 2 sampel masuk kelas D. Mayoritas kelas pada node ini adalah A.
- Cabang Kiri dari Node Root (Konsumsi Bahan Bakar <= 995000.0)**
Ketika konsumsi bahan bakar kurang dari atau sama dengan 995000.0, atribut berikutnya yang dipertimbangkan adalah "Negara" dengan batas 143.5. Mayoritas sampel (96 dari 97) masuk kelas A ketika negara kurang dari atau sama dengan 143.5. Jika negara lebih dari 143.5, hanya ada satu sampel yang masuk kelas B.
- Cabang Kanan dari Node Root (Konsumsi Bahan Bakar > 995000.0)**
Ketika konsumsi bahan bakar lebih dari 995000.0, atribut berikutnya adalah "Jumlah Kendaraan Lapis Baja Unit" dengan batas 162970.5. Jika jumlah kendaraan lapis baja kurang dari atau sama dengan 162970.5, semua sampel (17) masuk kelas C. Jika jumlah kendaraan lapis baja lebih dari 162970.5, kedua sampel masuk kelas D.

Pohon keputusan ini menunjukkan bahwa konsumsi bahan bakar dan jumlah kendaraan lapis baja merupakan faktor yang signifikan dalam memprediksi kelas yang berbeda terkait penggunaan bahan bakar dan kendaraan tempur militer pada suatu negara, dengan negara dan jumlah kendaraan lapis baja digunakan untuk mengklasifikasikan sampel lebih lanjut setelah menentukan konsumsi bahan bakar, serta mayoritas kelas pada setiap node membantu menentukan keputusan klasifikasi yang akurat sesuai dengan atribut yang diberikan.

5. KESIMPULAN DAN SARAN

Penggunaan algoritma C4.5 dalam memprediksi konsumsi bahan bakar kendaraan tempur militer memiliki potensi untuk meningkatkan efisiensi dan efektivitas pengelolaan sumber daya serta perencanaan strategi yang lebih efektif. Algoritma C4.5 terbukti membantu dalam analisis data dan pembangunan pohon keputusan dari data pelatihan. Dengan memahami pola penggunaan bahan bakar, suatu negara dapat mengoptimalkan alokasi sumber daya dan merencanakan strategi yang lebih efektif. Untuk melanjutkan penelitian ini perlu dikembangkan model prediksi yang lebih akurat dan mempertimbangkan faktor-faktor lain yang dapat memengaruhi konsumsi bahan bakar kendaraan tempur militer, seperti kondisi geografis, cuaca, dan jenis misi militer yang dilakukan.

DAFTAR PUSTAKA

- [1] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [2] L. Y. Lumban Gaol, M. Safii, and D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 97–106, 2021, doi: 10.30645/brahmana.v2i2.71.
- [3] I. Ubaedi and Y. M. Djaksana, "Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 17–26, 2022, doi: 10.30656/jsii.v9i1.3505.
- [4] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5," *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [5] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Januari*, vol. 2, no. 1, pp. 39–46, 2021.
- [6] D. P. S. Sinaga, Rini Marwati, and Bambang Avip Priatna Martadiputra, "Aplikasi Web Prediksi Dampak Gempa di Indonesia Menggunakan Metode Decision Tree dengan Algoritma C4.5," *JMT J. Mat. dan Terap.*, vol. 5, no. 2, pp. 97–108, 2023, doi: 10.21009/jmt.5.2.5.
- [7] R. Musfikar, H. Apriadinata, and B. Yusuf, "Aplikasi Prediksi Prestasi pada Siswa Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 13, no. 2, pp. 148–162, 2023, doi: 10.34010/jamika.v13i2.10649.
- [8] K. R. Dewi, K. F. Mauladi, and Masruroh, "Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri," *Semin. Nas. Inov. Teknol.*, vol. 25, pp. 109–114, 2020.
- [9] R. Dewi, Z. Hanif, and I. Santoso, "Analisa Kepuasan Konsumen Pada Laundry Menggunakan Algoritma C4.5," *Journals.Upi-Yai*, vol. 7, no. 2, pp. 133–141, 2023.
- [10] D. E. Yanti, L. Framesti, and A. Desiani, "Perbandingan Algoritma C4.5 Dan Svm Dalam Klasifikasi Penyakit Anemia," *J. Inform. Polinema*, vol. 9, no. 4, pp. 427–434, 2023, doi: 10.33795/jip.v9i4.1381.