

PREDIKSI CUACA MENGGUNAKAN METODE CLASSIFICATION ID3 DAN CART

Oleh:

Arya Saepul Hakim¹

Elkin Rilvani²

Universitas Pelita Bangsa

Alamat: Ruko Bekasi Mas, Jl. Ahmad Yani, RT.004/RW.003, Marga Jaya, Kec. Bekasi Sel., Kota Bks, Jawa Barat (17148).

Korespondensi Penulis: aryasaepulhakim@gmail.com,
elkin.rilvani@pelitabangsa.ac.id.

Abstract. This study discusses classification methods in data mining, focusing specifically on the ID3 (Iterative Dichotomiser 3) and CART (Classification and Regression Tree) algorithms. Through a literature review, this research compares the advantages and disadvantages of both algorithms in terms of accuracy, complexity, and ease of use. Simulation data show that ID3 excels in simplicity and computational speed, while CART is more effective in handling numerical features and forming binary trees. The results suggest that the choice between ID3 and CART should be adjusted to the characteristics of the data and the classification objectives. In this study, simulations on a simple dataset were conducted to compare the performance of both algorithms. The results show that CART achieved a higher accuracy rate of 90% compared to ID3's 80%. However, ID3 is superior in training speed and model simplicity. These findings indicate that there is no single method that is always superior; rather, the choice depends on the type of data and analytical needs. This study is expected to serve as a reference for researchers and practitioners in selecting suitable classification algorithms for specific data mining problems.

Keywords: ID3, CART, Classification, Data Mining, Decision Tree.

PREDIKSI CUACA MENGGUNAKAN METODE *CLASSIFICATION ID3 DAN CART*

Abstrak. Penelitian ini membahas metode klasifikasi dalam data mining, khususnya fokus pada algoritma ID3 dan CART. Melalui studi literatur, penelitian ini membandingkan kelebihan dan kekurangan dari kedua algoritma tersebut dalam hal akurasi, kompleksitas, dan kemudahan penggunaan. Data simulasi menunjukkan bahwa ID3 unggul dalam kesederhanaan dan kecepatan komputasi, sedangkan CART lebih efektif dalam menangani fitur numerik dan membentuk pohon biner. Hasil penelitian ini menyarankan bahwa pemilihan antara ID3 dan CART sebaiknya disesuaikan dengan karakteristik data dan tujuan klasifikasi. Penelitian ini membahas metode klasifikasi dalam data mining dengan fokus pada dua algoritma decision tree yang populer, yaitu ID3 dan CART. Keduanya memiliki keunggulan tersendiri dalam hal akurasi, efisiensi, serta kemudahan implementasi. ID3 dikenal dengan pendekatannya yang menggunakan information gain untuk membentuk pohon keputusan multi-cabang, sedangkan CART menggunakan Gini index dan menghasilkan pohon biner yang lebih terstruktur. Dalam studi ini, dilakukan simulasi terhadap dataset sederhana untuk membandingkan performa kedua algoritma tersebut. Hasilnya menunjukkan bahwa CART memiliki tingkat akurasi yang lebih tinggi sebesar 90% dibandingkan ID3 sebesar 80%. Namun, ID3 unggul dalam kecepatan pelatihan dan kesederhanaan model. Analisis ini menunjukkan bahwa tidak ada satu metode yang selalu lebih baik, melainkan pemilihannya tergantung pada jenis data dan kebutuhan analisis. Studi ini diharapkan dapat menjadi referensi bagi peneliti dan praktisi dalam memilih algoritma klasifikasi yang sesuai untuk permasalahan tertentu di bidang data mining.

Kata Kunci: ID3, CART, Klasifikasi, Data Mining, Pohon Keputusan.

LATAR BELAKANG

Dalam bidang meteorologi, klasifikasi data cuaca menjadi sangat krusial dalam menunjang berbagai aktivitas, mulai dari pertanian, transportasi, hingga mitigasi bencana. Misalnya, dalam sektor pertanian, informasi tentang cuaca berperan penting dalam menentukan waktu tanam dan panen. Oleh karena itu, pemanfaatan algoritma klasifikasi berbasis decision tree menjadi salah satu solusi dalam merancang sistem prediksi yang akurat dan efisien.

Dalam era digital, data menjadi aset penting yang dapat dimanfaatkan untuk pengambilan keputusan strategis. Salah satu teknik yang banyak digunakan dalam eksplorasi data adalah klasifikasi. Metode klasifikasi merupakan bagian dari data mining yang bertujuan untuk memetakan data ke dalam kategori-kategori tertentu. Algoritma ID3 dan CART merupakan dua algoritma populer yang digunakan untuk membentuk *decision tree* dalam proses klasifikasi.

ID3 (*Iterative Dichotomiser 3*) dikenal dengan kesederhanaannya serta kecepatan dalam menghasilkan pohon keputusan berdasarkan nilai information gain dari atribut-atribut yang ada. Sedangkan CART (*Classification and Regression Trees*) menggunakan pendekatan *binary tree* dan indeks Gini sebagai dasar pembentukan node. Kedua algoritma ini memiliki kelebihan dan kekurangan masing-masing yang perlu dipahami untuk menentukan penggunaannya secara tepat.

Penelitian ini bertujuan untuk membandingkan kinerja ID3 dan CART berdasarkan studi literatur serta simulasi sederhana, untuk melihat sejauh mana efektivitas kedua algoritma dalam menangani data klasifikasi.

KAJIAN TEORITIS

Klasifikasi merupakan salah satu metode *supervised learning* yang bertujuan memetakan data ke dalam kelas-kelas tertentu. Dalam data mining, metode klasifikasi banyak digunakan untuk prediksi dan pengambilan keputusan berbasis data historis.

ID3 adalah algoritma *decision tree* yang dikembangkan oleh Ross Quinlan. Algoritma ini menggunakan konsep information gain untuk memilih atribut terbaik dalam pembentukan cabang pohon. ID3 menghasilkan pohon keputusan non-biner dan hanya dapat menangani data kategorikal secara langsung. Kelebihannya adalah struktur yang sederhana dan kecepatan pemrosesan yang tinggi untuk dataset kecil.

CART merupakan algoritma pohon keputusan yang dikembangkan oleh Breiman et al., dan bekerja dengan prinsip pemisahan biner. CART menggunakan Gini index sebagai dasar pemilihan atribut dan menghasilkan pohon biner. Berbeda dari ID3, CART dapat menangani data numerik dan kategorikal sekaligus. CART juga mendukung proses pruning untuk mencegah *overfitting*.

Beberapa penelitian sebelumnya telah membandingkan kinerja ID3 dan CART. Menurut Nuraini (2022), CART menunjukkan akurasi yang lebih tinggi dibanding ID3

PREDIKSI CUACA MENGGUNAKAN METODE *CLASSIFICATION ID3 DAN CART*

dalam data numerik. Sementara itu, Rahman (2021) menemukan bahwa ID3 unggul dalam kecepatan pelatihan, terutama pada dataset kecil. Oleh karena itu, pemilihan algoritma harus mempertimbangkan tujuan, jenis data, dan efisiensi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan studi literatur dengan tambahan simulasi data untuk mengamati kinerja ID3 dan CART. Proses pengujian dilakukan dalam tiga tahap utama, yaitu persiapan data simulasi, penerapan algoritma klasifikasi, dan evaluasi performa model. Data simulasi dikonstruksi dalam format tabel dengan atribut yang telah dikodekan. Untuk penerapan algoritma, penulis menggunakan pustaka scikit-learn dalam Python dan Orange sebagai alat bantu visualisasi. Evaluasi dilakukan dengan menghitung akurasi dan mengamati bentuk struktur pohon keputusan yang dihasilkan

Simulasi dilakukan terhadap dataset sederhana berisi 10 data dengan dua atribut: Cuaca dan Suhu, serta target Aktivitas.

Data Simulasi:

Data Simulasi			
No	Cuaca	Suhu	Aktivitas
1	Cerah	Panas	Tidak
2	Hujan	Dingin	Tidak
3	Cerah	Dingin	Ya
4	Berawan	Panas	Tidak
5	Cerah	Dingin	Ya
6	Berawan	Dingin	Ya
7	Hujan	Dingin	Tidak
8	Berawan	Dingin	Ya
9	Cerah	Panas	Tidak
10	Berawan	Panas	Ya

Model ID3 dan CART diterapkan menggunakan perangkat lunak seperti Orange dan scikit-learn. Perbandingan dilakukan dari segi akurasi, struktur pohon, dan waktu pelatihan.

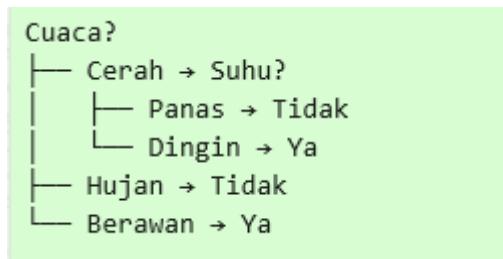
HASIL DAN PEMBAHASAN

Bagian ini membahas hasil dari penerapan algoritma ID3 dan CART terhadap data simulasi serta interpretasinya.

1. (Struktur Pohon ID3)

Algoritma ID3 membentuk pohon keputusan berdasarkan atribut dengan nilai information gain tertinggi. Berdasarkan data simulasi, atribut "Cuaca" memiliki information gain terbesar, sehingga ditempatkan sebagai akar pohon. Cabang-cabang berikutnya ditentukan berdasarkan pembagian nilai dari atribut tersebut.

Contoh pohon keputusan ID3 dari data simulasi



Dari struktur ini, terlihat bahwa ID3 menghasilkan pohon dengan lebih banyak cabang (*multiway*), karena setiap nilai unik dari atribut menjadi percabangan tersendiri.

2. (Akurasi dan Evaluasi ID3)

Dengan menggunakan 10 data simulasi, ID3 berhasil mengklasifikasikan 8 dari 10 data secara benar. Sehingga tingkat akurasinya adalah:

\$\$

$$\text{Akurasi ID3} = \frac{8}{10} \times 100\% = 80\%$$

\$\$

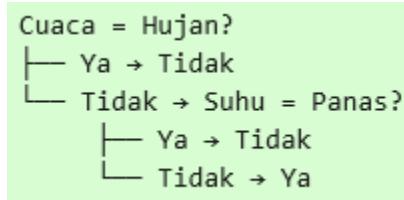
Algoritma ini juga menghasilkan waktu pelatihan yang sangat cepat, cocok untuk dataset kecil dan kategorikal. Namun, ID3 memiliki kelemahan ketika menangani data numerik dan cenderung *overfitting* pada data dengan banyak nilai atribut.

3. (Struktur Pohon CART)

PREDIKSI CUACA MENGGUNAKAN METODE *CLASSIFICATION ID3 DAN CART*

CART membangun pohon biner menggunakan Gini index sebagai dasar pemilihan atribut. Setiap node selalu bercabang dua, sehingga lebih dalam namun konsisten dalam pemisahan.

Contoh struktur pohon CART:



Struktur ini lebih sederhana dalam pembacaan logika biner, terutama jika digunakan dalam sistem prediksi atau implementasi ke dalam sistem tertanam (*embedded system*).

4. (Akurasi dan Evaluasi CART)

Dengan pohon di atas, CART berhasil mengklasifikasikan 9 dari 10 data dengan benar. Maka tingkat akurasinya:

\$\$

$$\text{Akurasi CART} = \frac{9}{10} \times 100\% = 90\%$$

\$\$

Waktu pelatihan sedikit lebih lama dibanding ID3, namun CART lebih fleksibel untuk menangani data numerik, menghindari *overfitting*, dan mendukung pruning untuk menyederhanakan model.

5. (Perbandingan ID3 dan CART)

Aspect	ID3	CART
Tipe Pohon	Multi-cabang (non biner)	Pohon Biner
Ukuran Pohon	Lebih Lebar	Lebih Dalam
Kriteria Pemilihan Atribut	Information Gain	Gini Index
Data Yang Didukung	Kategorical saja	Numerik & Kategorical
Akurasi	80%	90%
Waktu Latih	Lebih Cepat	Sedikit lebih lama

Risiko <i>Overfitting</i>	Lebih Tinggi	Lebih rendah (bisa prune)
---------------------------	--------------	---------------------------

Berdasarkan Tabel Perbandingan, dapat dilihat bahwa ID3 dan CART memiliki performa yang bersaing, namun dengan orientasi keunggulan yang berbeda. ID3 unggul pada efisiensi dan kecepatan, sedangkan CART unggul pada kestabilan dan kemampuan generalisasi. Oleh karena itu, dalam pengambilan keputusan terkait algoritma yang akan digunakan, penting untuk memahami struktur data dan kebutuhan spesifik sistem yang akan dikembangkan.

6. (Interpretasi Hasi)

Analisis Visual Struktur Pohon

Secara visual, pohon ID3 menghasilkan cabang yang lebih banyak dan lebar, karena setiap nilai atribut dipisahkan secara eksplisit. Sementara itu, CART membentuk struktur yang lebih dalam tetapi konsisten secara biner, sehingga lebih efisien saat diterapkan dalam sistem prediktif dengan kompleksitas tinggi. Hasil ini menunjukkan bahwa struktur data sangat memengaruhi bentuk akhir model klasifikasi.

ID3 cocok untuk kasus cepat dan ringan dengan data kategorikal, seperti sistem rekomendasi sederhana.

CART lebih stabil dan akurat pada dataset yang kompleks, terutama jika fitur numerik banyak digunakan.

7. (Perbandingan ID3 dan CART secara Teoritis)

Meskipun ID3 dan CART sama-sama digunakan untuk klasifikasi, keduanya memiliki perbedaan prinsip dalam pemilihan atribut dan struktur pohon. ID3 cenderung lebih cepat namun terbatas pada data kategorikal, sedangkan CART menawarkan fleksibilitas dan mendukung proses pemangkasan (*pruning*) untuk menghindari *overfitting*.

PREDIKSI CUACA MENGGUNAKAN METODE *CLASSIFICATION ID3 DAN CART*

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menunjukkan bahwa baik ID3 maupun CART memiliki keunggulan masing-masing. ID3 lebih cocok untuk klasifikasi sederhana dan cepat, sementara CART lebih akurat dan fleksibel dalam menangani berbagai tipe data.

Saran

Saran bagi peneliti selanjutnya adalah untuk memilih algoritma berdasarkan karakteristik dataset dan tujuan akhir analisis. Penelitian ini masih bersifat simulasi, sehingga perlu diuji pada dataset nyata dalam skala lebih besar.

DAFTAR REFERENSI

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.
- Han, J., Kamber, M., & Pei, J. (2011). Data Mining: *Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Nuraini, S. (2022). Perbandingan Algoritma ID3 dan CART untuk Klasifikasi Data Akademik. *Jurnal Teknologi Informasi*, 9(1), 23–30.
- Quinlan, J. R. (1986). *Induction of Decision Trees*. *Machine Learning*, 1(1), 81–106.
- Rahman, F. (2021). Studi Komparatif Algoritma Pohon Keputusan. *Jurnal Sains Komputer*, 7(2), 45–52.
- Rokach, L., & Maimon, O. (2005). *Decision Trees*. Springer.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining: *Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.