

Rangkuman Perkuliahan: Implementasi Algoritma Decision Tree ID3 dan C4.5 dalam Klasifikasi Data Diskrit dan Numerik

Muhammad Labib¹, Muhammad Shadaam Haidar Yuwono²

^{1,2} Program Studi Teknik Informatika, STMIK Tazkia Bogor
E-mail : 241552010005.shadaam@student.stmik.tazkia.ac.id

Abstrak

Decision Tree merupakan salah satu metode klasifikasi yang banyak digunakan dalam bidang data mining dan machine learning karena menghasilkan model yang mudah dipahami dan diinterpretasikan. Dua algoritma decision tree yang populer adalah ID3 dan C4.5. Algoritma ID3 digunakan untuk data diskrit, sedangkan C4.5 merupakan pengembangan yang mampu menangani data numerik dengan menggunakan threshold dan gain ratio. Paper ini membahas konsep dasar decision tree, proses perhitungan entropy, information gain, dan gain ratio, serta membandingkan penerapan algoritma ID3 dan C4.5 menggunakan Play Tennis Dataset dan Iris Dataset. Hasil pembahasan menunjukkan bahwa algoritma C4.5 memiliki fleksibilitas yang lebih baik dibandingkan ID3 dalam menangani data dunia nyata.

Kata kunci: Decision Tree, ID3, C4.5, Entropy, Gain Ratio

1. Pendahuluan

Perkembangan teknologi informasi menyebabkan meningkatnya jumlah data yang dihasilkan dalam berbagai bidang. Data tersebut perlu dianalisis untuk menghasilkan informasi yang bervalue dan dapat digunakan dalam proses pengambilan keputusan. Salah satu metode yang banyak digunakan dalam klasifikasi data adalah decision tree.

Decision tree merupakan metode pembelajaran mesin yang merepresentasikan proses pengambilan keputusan dalam bentuk struktur pohon. Algoritma ID3 merupakan salah satu algoritma decision tree awal yang menggunakan information gain sebagai dasar pemilihan atribut. Namun, algoritma ini memiliki keterbatasan karena hanya dapat digunakan pada data dengan atribut diskrit. Untuk mengatasi keterbatasan tersebut, dikembangkan algoritma C4.5 yang mampu menangani data numerik serta mengurangi bias pemilihan atribut.

2. Tinjauan Pustaka

2.1 Supervised Learning

Supervised learning adalah metode pembelajaran mesin yang menggunakan data berlabel. Setiap data memiliki atribut (fitur) dan target (label) yang menjadi acuan dalam proses pelatihan model..

2.2 Decision Tree

Decision tree adalah model prediksi yang merepresentasikan keputusan dalam bentuk struktur pohon yang terdiri dari root, node, cabang, dan leaf. Setiap node merepresentasikan atribut, sedangkan leaf merepresentasikan hasil klasifikasi.

2.3 Algoritma ID3

Algoritma ID3 membangun decision tree dengan cara memilih atribut yang memiliki information gain terbesar sebagai root atau node selanjutnya. Proses ini dilakukan secara rekursif hingga seluruh data berada pada kondisi murni.

2.4 Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian atau ketidakmurnian suatu dataset.

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

2.5 Information Gain

Information gain mengukur pengurangan entropy setelah dataset dibagi berdasarkan atribut tertentu.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in A} \frac{|S_v|}{|S|} \times \text{Entropy}(S_v)$$

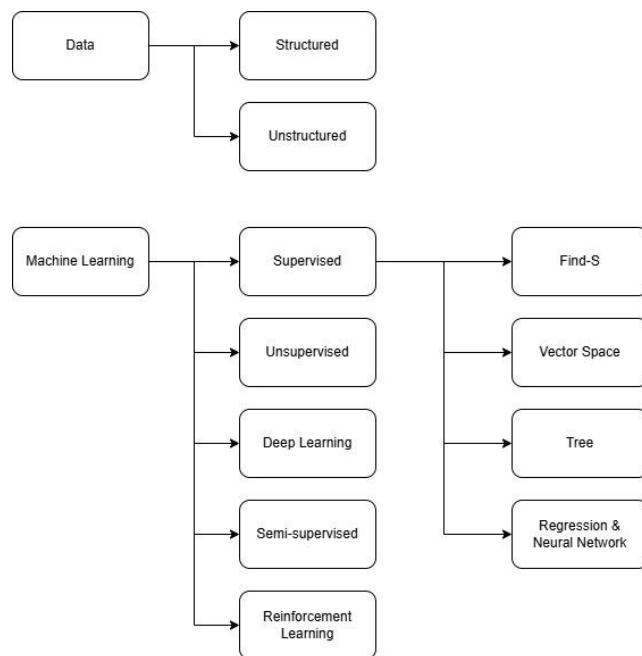
3. Hasil dan Pembahasan

Hasil pengujian pada Play Tennis Dataset menunjukkan bahwa atribut Outlook memiliki nilai information gain tertinggi sehingga dipilih sebagai node akar. Pada Iris Dataset, atribut Petal Length menghasilkan gain ratio tertinggi setelah dilakukan penentuan threshold. Hal ini menunjukkan bahwa algoritma C4.5 lebih fleksibel dibandingkan ID3 dalam menangani atribut numerik.

4. Kesimpulan

Hasil pengujian pada Play Tennis Dataset menunjukkan bahwa atribut Outlook memiliki nilai information gain tertinggi sehingga dipilih sebagai node akar. Pada Iris Dataset, atribut Petal Length menghasilkan gain ratio tertinggi setelah dilakukan penentuan threshold. Hal ini menunjukkan bahwa algoritma C4.5 lebih fleksibel dibandingkan ID3 dalam menangani atribut numerik.

5. Kesimpulan



Daftar Pustaka

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2012.
- [2] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [3] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.