

## Rangkuman Perkuliahan: Implementasi Algoritma Decision Tree ID3 Pada Supervised Learning

Muhammad Labib<sup>1</sup>, Muhammad Shadaam Haidar Yuwono<sup>2</sup>

<sup>1,2</sup> Program Studi Teknik Informatika, STMIK Tazkia Bogor  
E-mail : [241552010005.shadaam@student.stmik.tazkia.ac.id](mailto:241552010005.shadaam@student.stmik.tazkia.ac.id)

### Abstrak

Machine Learning merupakan cabang kecerdasan buatan yang memungkinkan sistem komputer belajar dari data tanpa diprogram secara eksplisit. Salah satu pendekatan utama dalam Machine Learning adalah Supervised Learning, di mana data latih telah memiliki label kelas. Decision Tree merupakan algoritma Supervised Learning yang populer karena mudah dipahami dan diinterpretasikan. Penelitian ini membahas algoritma Decision Tree dengan fokus pada metode ID3 (Iterative Dichotomiser 3), termasuk konsep entropy dan information gain sebagai dasar pemilihan atribut terbaik. Selain itu, diperkenalkan algoritma C4.5 sebagai pengembangan dari ID3 yang mampu menangani data kontinu dan nilai yang hilang. Hasil pembahasan menunjukkan bahwa ID3 efektif untuk data diskrit, namun memiliki keterbatasan yang dapat diatasi oleh C4.5.

**Kata kunci :** Machine Learning, Supervised Learning, Decision Tree, ID3, Entropy, Information Gain

### 1. Pendahuluan

Machine Learning merupakan salah satu bidang dalam kecerdasan buatan yang berkembang pesat dan banyak diterapkan dalam berbagai bidang, seperti klasifikasi, prediksi, dan sistem pendukung keputusan. Machine Learning memungkinkan sistem komputer mempelajari pola dari data dan menghasilkan keputusan secara otomatis berdasarkan pengalaman sebelumnya.

Salah satu pendekatan utama dalam Machine Learning adalah Supervised Learning, yaitu metode pembelajaran yang menggunakan data berlabel. Dalam pendekatan ini, model dilatih menggunakan pasangan data input dan output sehingga mampu memprediksi kelas dari data baru. Algoritma Decision Tree merupakan salah satu metode Supervised Learning yang banyak digunakan karena strukturnya yang sederhana dan mudah dipahami.

### 2. Supervised Learning

Supervised Learning adalah metode pembelajaran mesin yang menggunakan dataset berlabel, di mana setiap data terdiri dari atribut dan target kelas. Tujuan utama dari metode ini adalah membangun model yang dapat melakukan prediksi atau klasifikasi secara akurat.

Beberapa algoritma yang termasuk dalam Supervised Learning antara lain Decision Tree, Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine. Pada penelitian ini, fokus pembahasan diarahkan pada algoritma Decision Tree.

### 3. Decision Tree

Decision Tree merupakan model prediksi berbentuk struktur pohon yang terdiri dari node, cabang, dan daun. Node merepresentasikan atribut, cabang merepresentasikan nilai atribut, dan daun merepresentasikan keputusan atau kelas. Algoritma ini bekerja dengan membagi dataset menjadi beberapa subset berdasarkan atribut tertentu hingga mencapai kondisi yang homogen.

#### 4. ID3

ID3 (Iterative Dichotomiser 3) merupakan algoritma pembentukan Decision Tree yang menggunakan konsep entropy dan information gain untuk menentukan atribut terbaik sebagai node.

##### 4.2 Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian dalam suatu dataset. Semakin kecil nilai entropy, semakin homogen data tersebut. Entropy dirumuskan sebagai berikut:

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

##### 4.3 Information Gain

Information Gain digunakan untuk mengukur kemampuan suatu atribut dalam membagi data. Atribut dengan nilai information gain terbesar akan dipilih sebagai node utama. Rumus information gain ditunjukkan pada persamaan berikut:

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

#### 5. Keterbatasan Algoritma ID3

Algoritma ID3 memiliki beberapa keterbatasan, antara lain tidak mampu menangani data kontinu secara langsung, tidak dapat mengelola nilai yang hilang (missing value), serta rentan terhadap overfitting pada dataset tertentu.

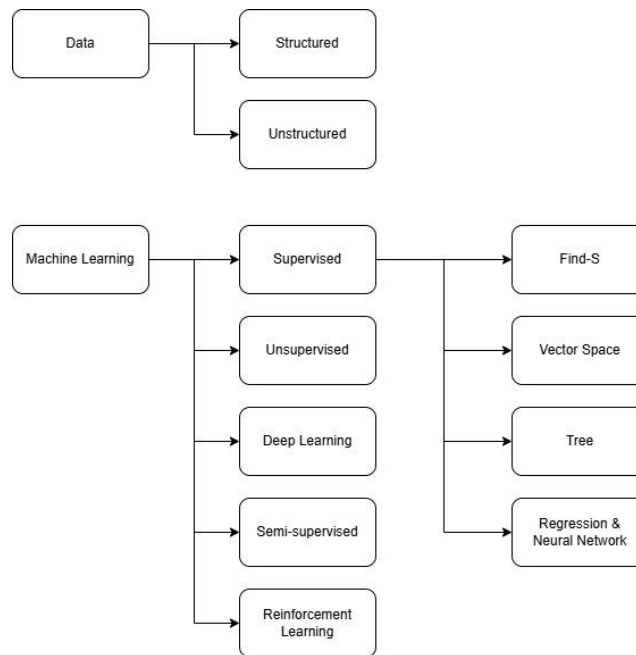
#### 6. Algoritma C4.5

C4.5 merupakan pengembangan dari algoritma ID3 yang dirancang untuk mengatasi keterbatasan tersebut. Algoritma ini mampu menangani data kontinu, missing value, serta menggunakan Gain Ratio sebagai kriteria pemilihan atribut sehingga menghasilkan pohon keputusan yang lebih optimal.

#### 7. Kesimpulan

Decision Tree merupakan algoritma Supervised Learning yang efektif dan mudah dipahami. Algoritma ID3 menggunakan entropy dan information gain sebagai dasar pembentukan pohon keputusan, namun memiliki keterbatasan dalam menangani data kontinu dan missing value. Algoritma C4.5 hadir sebagai pengembangan yang mampu mengatasi keterbatasan tersebut dan menghasilkan model yang lebih stabil dan akurat.

## 8. Mindmap



## 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.