

Rangkuman Perkuliahan: Implementasi Algoritma ID3 dalam Pembentukan Decision Tree untuk Klasifikasi Data

Muhammad Labib¹, Muhammad Shadaam Haidar Yuwono²

^{1,2} Program Studi Teknik Informatika, STMIK Tazkia Bogor
E-mail : 24155201005.shadaam@student.stmik.tazkia.ac.id

Abstrak

Algoritma ID3 (Iterative Dichotomiser 3) merupakan salah satu algoritma dasar dalam machine learning yang digunakan untuk membangun model *decision tree* pada permasalahan klasifikasi. Algoritma ini bekerja dengan memilih atribut yang memiliki nilai *information gain* tertinggi sebagai akar pohon keputusan. Penelitian ini bertujuan untuk menjelaskan konsep, tahapan, serta implementasi algoritma ID3 menggunakan studi kasus dataset *Play Tennis*. Hasil dari penelitian ini menunjukkan bahwa algoritma ID3 mampu membentuk aturan keputusan yang sistematis dan mudah dipahami berdasarkan perhitungan entropi dan information gain.

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

1. Pendahuluan

Perkembangan teknologi informasi mendorong pemanfaatan data dalam pengambilan keputusan. Salah satu pendekatan yang banyak digunakan adalah *machine learning*, khususnya *supervised learning* untuk permasalahan klasifikasi. Decision tree merupakan metode klasifikasi yang populer karena hasilnya mudah dipahami dan dapat direpresentasikan dalam bentuk struktur pohon.

Algoritma ID3 diperkenalkan oleh J. Ross Quinlan sebagai metode untuk membangun decision tree berdasarkan konsep teori informasi. ID3 menggunakan ukuran ketidakmurnian data yang disebut *entropy* serta ukuran kontribusi atribut yang disebut *information gain* untuk menentukan struktur pohon keputusan.

2. Landasan Teori

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.

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

3. Kesimpulan

Penelitian ini menggunakan dataset *Play Tennis* yang terdiri dari 14 data dengan empat atribut, yaitu Outlook, Temperature, Humidity, dan Wind, serta satu label Play Tennis (Yes/No). Tahapan penelitian meliputi:

1. Menghitung entropy total dataset.
2. Menghitung entropy untuk setiap nilai atribut.
3. Menghitung information gain setiap atribut.
4. Menentukan atribut dengan gain tertinggi sebagai root.
5. Mengulangi proses hingga terbentuk decision tree lengkap.

4. Hasil dan Pembahasan

Berdasarkan perhitungan, diperoleh entropy total dataset sebesar 0,94. Nilai information gain dari setiap atribut adalah sebagai berikut:

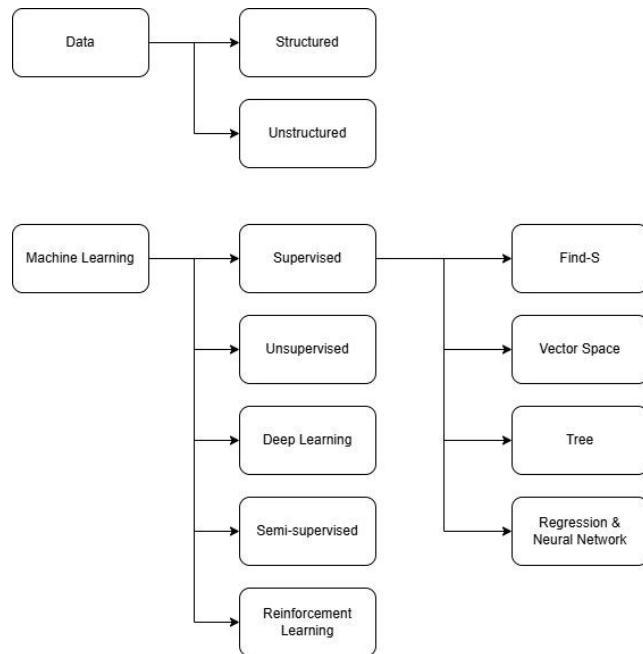
- Outlook: 0,246
- Temperature: 0,029
- Humidity: 0,151
- Wind: 0,048

Atribut Outlook memiliki nilai information gain tertinggi sehingga dipilih sebagai root decision tree. Cabang Overcast langsung menghasilkan keputusan Yes karena seluruh data bernilai positif. Untuk cabang Sunny, atribut Humidity dipilih sebagai node berikutnya, sedangkan untuk cabang Rain digunakan atribut Wind. Hasil akhir berupa aturan keputusan yang dapat digunakan untuk memprediksi apakah seseorang akan bermain tenis atau tidak.

5. Kesimpulan

Algoritma ID3 mampu membangun decision tree yang efektif dengan memanfaatkan perhitungan entropy dan information gain. Model yang dihasilkan bersifat interpretatif dan mudah dipahami, sehingga cocok digunakan untuk permasalahan klasifikasi dengan data kategorikal. Dataset Play Tennis membuktikan bahwa ID3 dapat menghasilkan aturan keputusan yang jelas dan sistematis.

6. 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.
- [3] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.