# Algoritma Iterative Dichotomizer 3 (ID3) Pengambilan Keputusan

## Decision Making Using Iterative Dichotomizer 3 Algorithm

Arief Kelik Nugroho[1], Dadang Iskandar[2]

[1]ariefkeliknugroho@gmail.com

[2]dadangiskandar83@gmail.com

*Teknik Informatika, Fakultas Teknik - Universitas Jenderal Soedirman*

*Abstrak* — **Tujuan makalah ini adalah menerapkan metode yang dapat digunakan dalam proses klasifikasi untuk memberikan keputusan. Model klasifikasi dapat digambarkan dalam berbagai bentuk, salah satunya adalah dengan menggunakan konsep pohon (*tree*). Pohon (*tree*) merupakan salah satu konsep teori graf yang paling penting. Pemanfaatan pohon dalam kehidupan sehari-hari adalah untuk menggambarkan hierarki dan memodelkan persoalan. *Iterative dichotomiser* 3 ( ID3 ) merupakan suatu metode dalam learning yang akan membangun sebuah pohon keputusan untuk mencari solusi dari persoalan. Pohon keputusan yang dihasilkan dengan menggunakan proses pencarian nilai terbaik (*the best classifier)* akan dijadikan seagai akar (*root*). Dalam penelitian ini akan dibahas model klasifikasi menggunakan Decision Tree dengan algoritma *Interactive Dichotomicer* 3 (ID3) berdasarkan pada kasus**

*Kata kunci* — **Pohon keputusan, *Interactive Dichotomicer* 3, Kasus**

*Abstract* — **The aim of this paper is apllying method that can used in a classification process for the decision making. A classification model can be describes in various form, one of which is using tree concept. Tree is one of the graph concept which is most important. Tree utilization in life is to describe and model the hierarchy problems. *Iterative Dichotomiser* 3 ( ID3 ) a method of learning that will build a decision tree to find solutions of the problems. The resulting decision tree by using the process is finding the best value (the best classifier) will serve as the root (root). In this paper discussed classification model a Decision Tree by using Dichotomicer Interactive 3 (ID3) algorithm based on case**

*Keyword* — ***Decision Tree*, *Interactive Dichotomicer* 3, *Case***

### PENDAHULUAN

Mahasiswa merupakan aset perguruan tinggi yang sangat penting. Dalam proses pengajaran, mahasiswa dituntut mampu berprestasi secara akademik maupun non akademik. Prestasi merupakan unsur penting dalam merupakan unsur penting dalam akademik. Kriteria yang dapat digunakan dalam penilaian dalam menetukan mahasiswa berprestasi dapat dilakukan dengan menentukan atribut-atribut, atribut tersebut diantranya adalah nilai indeks prestasi, psikologi atau kepribadian dan kemampuan mahasiswa dalam menganalisa, menjelaskan kajian teori dalam perkuliahan.

### TEORI

Penalaran berbasis kasus (*Case-Base Reasoning*) merupakan sebuah sistem yang menggunakan pengalaman lama untuk dapat mengerti dan menyelesaikan masalah baru. (Sankar dan Simon, 2004). Salah satu metode yang dapat digunakan adalah algoritma *ID3*. Algoritma ID3 dapat diimplementasikan menggunakan fungsi *rekursif* (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun *decision tree* (pohon keputusan) secara *top-down* (dari atas ke bawah), dimulai dengan atribut mana yang pertama kali harus dicek dan diletakkan sebagai *root*. Caranya adalah dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistic (yang banyak digunakan adalah *information gain*) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data.

#### Entropy dan Information Gain

Sebuah obyek yang diklasifikasikan dalam pohon harus dites nilai entropinya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impuryt* ,dan *homogenity* dari kumpulan data. Dari nilai entropy tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut.

$$\text{Entropy}(S) = - p+ \log2p + -p -\log2p \quad (1)$$

dimana :

- S : ruang (data) sample yang digunakan untuk training.
- p+ : jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.
- p- : jumlah yang bersolusi negatif (tidak mendukung) pada data sample untuk criteria tertentu.

Dari persamaan (1) dapat disimpulkan bahwa definisi entropy (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas.

Panjang kode untuk menyatakan informasi secara optimal adalah –log2p bits untuk messages yang mempunyai probabilitas p. Sehingga jumlah bit yang diperkirakan untuk mengekstraksi dalam kelas adalah :

- p+log2 p+ - p- log2 p-

Untuk suatu kumpulan data, maka dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut information gain, Secara matematis, infomation gain dari suatu atribut A,dituliskan sebagai berikut :

$$\text{Information Gain} = \sum_{\boldsymbol{Value(A)}} \frac{|\boldsymbol{Sv}|}{|\boldsymbol{S}|} \qquad (2)$$

$$\text{Gain}(S,A) = \text{Entropy}(S) - \text{Entropy}(Sv), \qquad (3)$$

dimana :

- A : atribut
- V : suatu nilai yang mungkin untuk atribut A
- Values (A) : himp yang mungkin untuk atribut A
- |Sv| : jum sampel untuk nilai v
- |S| : jum seluruh sampel data
- Entropy(Sv) : entropy untuk sampel-sampel yang memilki nilai v

## HASIL DAN PEMBAHASAN

Data uji 15 sampel kasus yang terdiri atas 3 atribut penilaian seperti pada tabel 1. Sedangkan 15 sampel kasus tersaji pada table 2.

Tabel 1. Parameter (Atribut) dari Setiap Kasus

| No | Parameter / Atribut | Kategori |
|---|---|---|
| 1 | IPK | Bagus,Cukup,Kurang |
| 2 | Psikologi/Kepribadian | Tinggi, Sedang, Rendah |
| 3 | Wawancara | Baik,Cukup |

Tabel 2. Tabel Kasus

| Mhs | Ipk | Psikologi/ Kepribadian | Wawancara | Diterima |
|---|---|---|---|---|
| P1 | Bagus | Tinggi | Baik | Ya |
| P2 | Bagus | Sedang | Baik | Ya |
| P3 | Bagus | Sedang | Cukup | Ya |
| P4 | Bagus | Rendah | Cukup | Tidak |
| P5 | Cukup | Tinggi | Baik | Ya |
| P6 | Cukup | Sedang | Baik | Ya |
| P7 | Cukup | Sedang | Cukup | Ya |
| P8 | Cukup | Rendah | Cukup | Tidak |
| P9 | Kurang | Tinggi | Baik | Ya |
| P10 | Kurang | Sedang | Cukup | Tidak |
| P11 | Kurang | Rendah | Baik | Ya |
| P12 | Bagus | Rendah | Baik | Tidak |
| P13 | Bagus | Rendah | Cukup | Tidak |
| P14 | Cukup | Rendah | Baik | Tidak |
| P15 | Cukup | Rendah | Cukup | Tidak |

Untuk analisis information gain, pada data tabel 2, jumlah kelas adalah 2, yaitu : ‘ya’ dan ‘tidak’ . Jumlah sampel C1 (‘ya’) adalah 8 dan jumlah sampel C2 (‘tidak’) adalah 7 .

C1 = Ya (p+)
C2 = Tidak(p-)
Jumlah sampel C1 = 8
Jumlah sampel C2 = 7

Nilai entropy untuk kumpulan sampel data S adalah :

Entropy (S) = - ( 8/15) log2 ( 8/15) – (7/15) log2 (7/15)
= 0.99679163198

Nilai Entropy pada masing-masing atribut dapat dihitung seperti pembahasan di bawah ini.

*A. Atribut IPK*

Tabel 3 Atribut IPK

| IPK | Diterima | Jumlah |
|---|---|---|
| Bagus | Ya | 3+ |
| Bagus | Tidak | 3- |
| Cukup | Ya | 3+ |
| Cukup | Tidak | 3- |
| Kurang | Ya | 2+ |
| Kurang | Tidak | 1- |

Nilai entropy dan *Information Gain* untuk IPK dapat dihitung seperti di bawah ini.

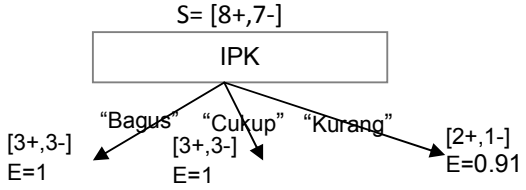
Entropy ($S_{Banyak}$) = - ( 3/6) log2 ( 3/6) – (3/6) log2 (3/6)
= 1

Entropy ($S_{Cukup}$) = - ( 3/6) log2 ( 3/6) – (3/6) log2 (3/6)
= 1

Entropy ($S_{Kurang}$)= - ( 2/3) log2 ( 2/3) – (1/3) log2 (1/3)
= 0.91829583405

Entropy IPK = 6/15 * Entropy ($S_{Banyak}$) + 6/15 * Entropy ($S_{Cukup}$) + 3/15 * Entropy ($S_{kurang}$)
= 6/15*1+ 6/15 *1+ 3/15* 0.91829583405
= 0.98365916681

IG (S,IPK) = Entropy(S) – Entropy IPK
= 0.99679163198– 0.98365916681
= 0.01313246517




Gambar 1. Information gain untuk IPK

*B. Atribut Psikologi/Kepribadian*

Tabel 4 Atribut Psikologi/Kepribadian

| Psikologi/ Kepribadian | Diterima | Jumlah |
|---|---|---|
| Tinggi | Ya | 3+ |
| Tinggi | Tidak | 0- |
| Sedang | Ya | 4+ |
| Sedang | Tidak | 1- |
| Rendah | Ya | 1+ |
| Rendah | Tidak | 6- |

Maka nilai entropy dan *Information Gain* untuk Psikologi/Kepribadian adalah :

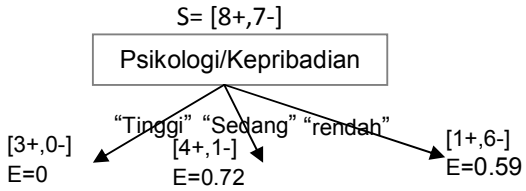
Entropy ($S_{tinggi}$) = - ( 3/3) log2 (3/3) – (0/3) log2 (0/3)
= 0

Entropy ($S_{sedangl}$) = - ( 4/5) log2 (4/5) – (1/5) log2 (1/5)
= 0.72192809488

Entropy ($S_{Rendahl}$) = - (1/7) log2 (1/7) – (6/7) log2 (6/7)
= 0.59167277858

Entropy Psikologi/Kepribadian
= 3/15 * Entropy ($S_{tinggi}$) + 5/15 * Entropy ($S_{sedang}$) + 5/15 * Entropy ($S_{rendah}$)
=5/15*0+5/15*0.72192809488+ 7/15*0.59167277858
= 0.51675666163

IG (S, Psikologi/Kepribadian) = Entropy(S) – Entropy Psikologi/Kepribadian
= 0.99679163198– 0.51675666163
= 0.48003497035




Gambar 2. Information gain psikologi/kepribadian

*C. Atribut Wawancara*

Tabel 5. wawancara

| Wawancara | Diterima | Jumlah |
|---|---|---|
| Baik | Ya | 6+ |
| Baik | Tidak | 2- |
| Cukup | Ya | 2+ |
| Cukup | Tidak | 5- |

Maka nilai entropy dan *Information Gain* Untuk Rasa adalah :

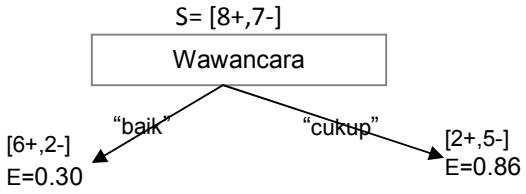
Entropy ($S_{Baik}$) = - (6/8) log2 (6/5) – (2/8) log2 (2/8)
= 0.30272419562

Entropy ($S_{cukup}$) = - (2/7) log2 ( 2/7) – (5/7) log2 (5/7)
= 0.86312056856

Entropy Wawancara
= 8/15*Entropy ($S_{Baik}$)+ 7/15 * Entropy ($S_{Cukup}$)
= 8/15*0.30272419562+7/15* 0.86312056856
= 0.56424250299

IG (S,wawancara) = Entropy(S) – Entropy wawancara
=0.99679163198– 0.56424250299
=0.43254912899

Berdasarkan hasil *Information Gain*, maka nilai IG (S, Psikologi/Kepribadian) adalah atribut yang memiliki IG terbesar dan menjadi node awal (root) yakni sebesar 0.48003497035




Gambar 3. Information gain wawancara

**Rekursi Level 0 iterasi ke-1**

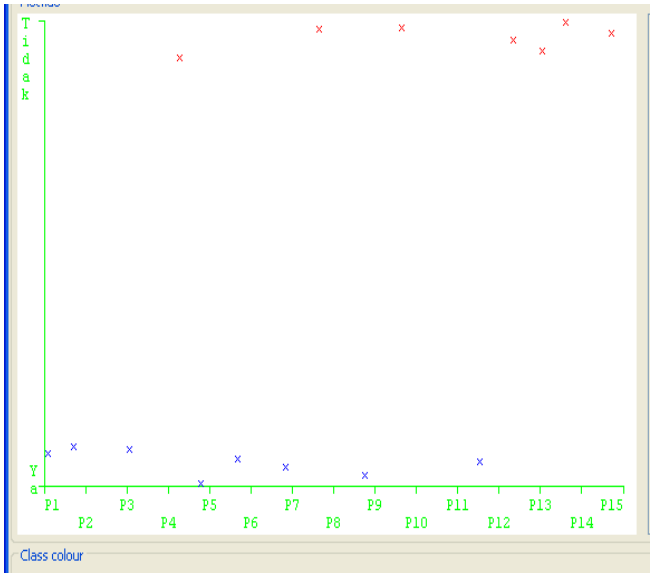
Memanggil sampel [8+,7-] dengan atribut target adalah =”Ya”. Hasil penghitungan nilai terbesar *information gain* (IG) yang akan dijadikan root adalah atribut psikologi/kepribadian dengan nilai IG= 0.48003497035, jumlah sampel adalah [8+,7-].

**Rekursi Level 1 iterasi ke-1**

Memanggil fungsi ID3 dengan kumpulan sampel berupa sampel baik [8+, 7-] atribut target = ‘Ya’ dan kumpulan atribut (wawancara,Psikologi/kepribadian). Semua sampel baik termasuk dalam kelas “ya” maka fungsi ini akan berhenti dan mengembalikan satu simpul tunggal Root dengan label ‘ya’.

**Rekursi level 0 , Itersi ke 2**

Pada proses rekursi level 0 iterasi ke 1, sudah dilakukan pengecekan untuk atribut ‘Psikotest’ dengan Wawancara ‘baik’. Untuk wawancara ‘cukup’, terdapat 7 sampel, berarti Sampel buruk tidak kosong. Sehingga, perlu memanggil fungsi ID3 dengan kumpulan sampel berupa Sampel buruk = [ 8+, 7- ]. AtributTarget =’Ya’, dan Kumpulan Atribut = { wawancara, IPK }. Proses itersai selesai samapai dengan pengecekan semua atribut dalam kasus.

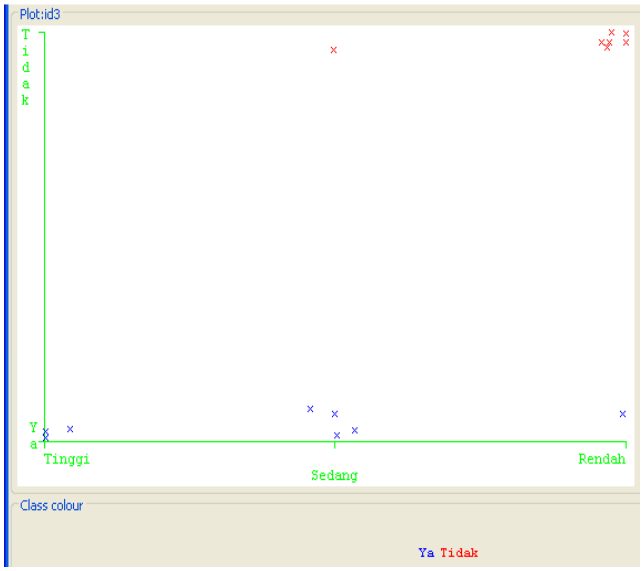



Gambar 4. Jumlah data dengan keputusan diterima "ya" atau "tidak"

Klasifikasi jumlah data berdasarkan ke 3 atribut di atas dengan melihat keputusan diterima "Ya" dan keputusan "Tidak" dengan jumlah sampel data adalah 15. Klasifikasi data berdasarkan atribut IPK dengan keputusan diterima "ya" dan "tidak" dengan data 15.

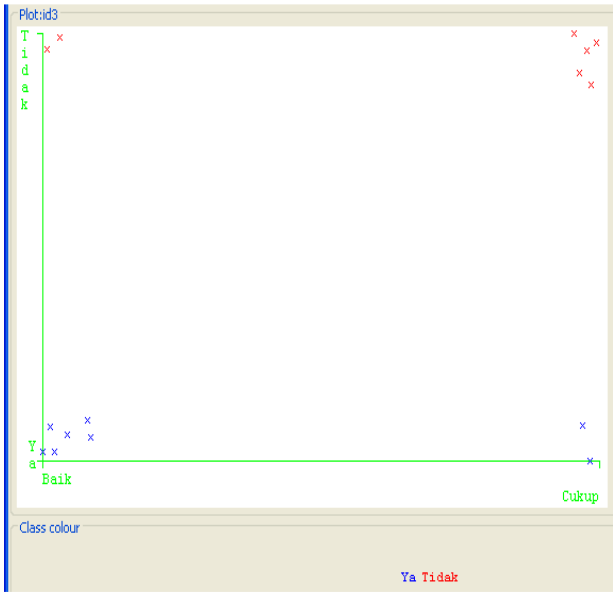
Gambar 5. Plot atribut ipk dengan keputusan diterima"ya" atau "tidak"

Klasifikasi jumlah data dengan atribut psikologi/kepribadian adalah sebagai berikut : atribut psikologi/kepribadian (tinggi) dengan keputusan diterima "ya"dan "tidak", psikologi/kepribadian (sedang) "ya" dan "tidak", psikologi/kepribadian (rendah) "ya" dan "tidak" sesuai dengan gambar 7.




Gambar 6. Plot atribut Psikologi/kepribadian dengan keputusan diterima "ya" atau "tidak"

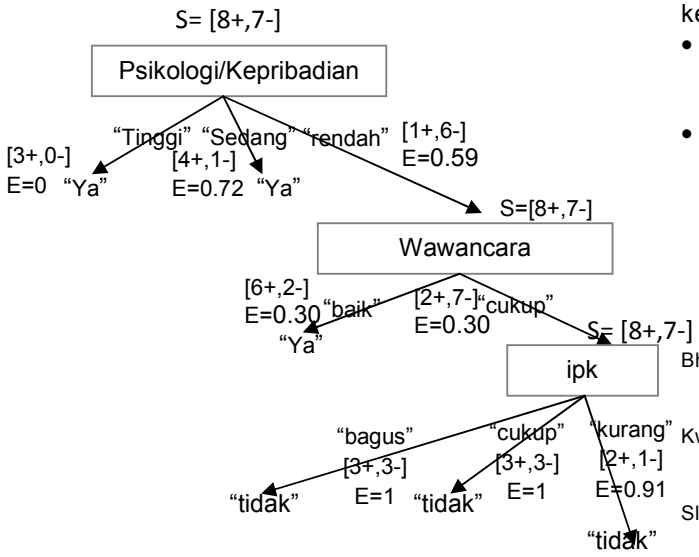
Klasifikasi data dengan atribut wawancara (baik) dengan keputusan "ya"dan "tidak" dan wawancara (cukup) "ya" dan "tidak" dijelaskan pada gambar 8.




Gambar 7. Plot wawancara dengan keputusan diterima "ya" atau "tidak"

**Pohon Keputusan**

Penelusuran mulai dari simpul Root menuju ketiga leaf node tersebut. Lakukan operasi *conjunction* ( ^ ) pada setiap simpul yang dikunjungi sampai ditemukan leaf node 'Ya'. Kemudian, lakukan operasi disjunction ( v ) pada penelusuran tersebut.




Gambar 8. Pohon keputusan

## KESIMPULAN

Penerapan algoritma ID3 terdapat konsep pada penanganan kasus yaitu Entropy (*measurement of uncertainty*) dan *Information Gain* (*measurement of purity*). Pengambilan keputusan dapat dilakukan dengan berbagai cara, salah satunya adalah penerapan Algoritma ID3.

Dari kasus-kasus dalam penelitian ini diperoleh kesimpulan seperti di bawah ini.

- Merujuk pada hasil penghitungan, nilai *Information Gain* yang menjadi root kasus dalam tabel 2 adalah psikologi/kepribadian.
- Dengan adanya pohon keputusan, dapat dirunut untuk kasus baru di luar kasus (15 data yang disajikan)

## DAFTAR PUSTAKA

Bhardwaj, R.,Vatta,S.,2013, Implementation of ID3 Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, Issue 6,Volume 3.

Kwon,B,O.,Sadeh,N.,2003,Applying Case-Based Reasoning And Multi-Agent Intelligent, School of Computer Science, Carnegie Mellon University

Slocum,M.,2012, Decision Making Using Id3 Algorithm, InSight: Rivier Academic Journal, Volume 8, Number 2

Tyasti,A,E.,Ispriyanti,D.,Hoyyi,A., 2015, Algoritma Iterative Dichotomiser 3 (Id3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta), JURNAL GAUSSIAN, Volume 4, Nomor 2

Watson,I.,1997,Applying Case-Based Reasoning:Techniques For Enterprise System,Morgan Kaufmann Publisher,U.K

Peng, W., Chen, J., Zhou, H. An Implementation of ID3 - Decision Tree Learning Algorithm. Retreived May 26, 2014, from http://cis.k.hosei.ac.jp/~rhuang/Miccl/AI-2/L10-src/DecisionTree2.pdf