

DECISION TREE MENGGUNAKAN ALGORITMA ID3 UNTUK MELAKUKAN DETEKSI PENYAKIT KANKER PAYUDARA

Putri Anisya Maulida Mutmainah

Jurusan Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Dian Nuswantoro

Jalan Nakula I No. 5 – 11 Semarang Jawa Tengah 50131, (024) 3517261

E-mail : 111201105835@mhs.dinus.ac.id , sekretariat@dinus.ac.id

Abstrak

Kanker Payudara merupakan penyakit dengan tingkat kematian yang tinggi, terutama bagi perempuan. Penyakit ini membutuhkan penanganan yang sesegera mungkin untuk dapat menekan angka kematian pada pasien. Pemeriksaan secara dini dan cepat perlu dilakukan dan diketahui oleh rumah sakit untuk mempercepat penanganan pada pasien Kanker Payudara. Iterative Dichotomizer (ID3) merupakan salah satu algoritma Klasifikasi yang digunakan pada Data Mining untuk melakukan pengambilan keputusan dari beberapa atribut yang telah ditentukan sesuai dengan Data yang diperoleh serta kebutuhan dari penggunanya. Algoritma ini biasanya menghitung nilai Entropy dari setiap atribut dan nilai Information Gain dari setiap atribut untuk dapat membentuk sebuah Decision Tree yang akan membantu dalam pengambilan keputusan dalam suatu permasalahan dari data yang di olah. Dalam penelitian ini, dilakukan penghitungan nilai Entropy pada setiap atribut yang berasal dari nilai hasil uji laboratorium pasien terduga Kanker Payudara. Dimana setiap atribut berpengaruh terhadap diagnosa Negatif atau Positif dari dokter yang bersangkutan. Selanjutnya, dihitung nilai Information Gain dari setiap atribut tersebut lalu pembentukan Pohon Keputusan berdasarkan dari atribut yang memiliki nilai Information Gain tertinggi hingga yang terendah. Dari hasil pengujian yang dilakukan, metode yang diusulkan oleh penulis dapat digunakan untuk membantu diagnosa pasien terduga Kanker Payudara lebih cepat dan tingkat keakuratan yang tinggi.

Kata Kunci: *Data Mining, Metode Klasifikasi, Iterative Dichotomizer 3, Kanker Payudara.*

Abstract

Breast cancer is a disease with a high mortality rate, especially for women. This disease requires treatment as soon as possible to be able to surpress the death rate in patients. Early and rapid examination needs to be done and known by the hospital to speed up the handling of in patients of breast cancer. Dichotomizer Iterative (ID3) is one classification algorithms used in Data Mining to do the decisions of some of the attributes that have been specified in accordance with the data acquired as well as the needs of its users. This algorithms usually calculate the Entropy values of any attributes and the value of Information Gain of each attribute to be able to form a Decision Tree that will

help in the decision making in a problem of data processed. In this study, performed the calculation value of Entropy in each attribute that is derived from the value of the patient's lab test result unexpected breast cancer. Where each of the attributes effect on Negative or Positive diagnosis of doctor in question. Furthermore, the calculated value of Information Gain of each of these attributes, then the establishment of decision tree based on the attribute that has the value of Information Gain is the highest to the lowest. From the results of testing performed, the method proposed by the author can be used to help diagnose a patient's unexpected breast cancer more quickly and to a high level of accuracy.

Keywords: Data Mining, Classification Method, Iterative Dichotomizer 3, Breast Cancer.

1.PENDAHULUAN

Penyakit kanker yang saat ini merupakan salah satu penyakit yang menyebabkan tingkat kematian yang tinggi. Terutama penyakit Kanker Payudara yang sangat rentan terjadi pada Wanita namun, juga dapat terjadi pada Pria. Tingkat kematian yang tinggi pada penyakit ini, disebabkan karena kurangnya pengetahuan masyarakat tentang gejala umum yang muncul pada penyakit kanker serta penderita kanker umumnya tidak pernah memeriksakan kondisinya lebih dini. Sehingga, saat terdiagnosa mengidap penyakit kanker penderita tersebut biasanya telah mencapai fase yang cukup ganas atau bahkan sudah sangat ganas. Hal ini, menyebabkan pula pengobatan dan recovery yang dilakukan oleh tenaga medis biasanya kurang maksimal. Karena sel kanker yang berada dalam tubuh sudah telanjur menyebar saat terdiagnosa. Faktanya, untuk mendeteksi penyakit kanker sejak dini dapat dilakukan dengan beberapa cara yang sederhana hingga ke pemeriksaan Tumor Marker melalui media Darah.

Dalam Tumor Marker yang di uji, terdapat tiga atribut utama yang dapat mempengaruhi diagnosa pasien yaitu CEA, Ca – 15 – 3, dan Ca.27.29. metode yang digunakan adalah Iterative

Dichotomizer (ID3) yang akan mempengaruhi dalam pengambilan keputusan dari pihak rumah sakit, dokter, maupun pasien untuk pengobatan lanjutan yang akan dilakukan untuk kesembuhan pasien itu sendiri.

2. METODE

Metode menggunakan Algoritma Klasifikasi Data Mining yaitu Algoritma Iterative Dichotomizer 3 (ID3). Algoritma ini merupakan algoritma klasifikasi pertama yang memiliki tingkat akurasi yang tinggi.

2.1 ALGORITMA ID3

Dalam penelitian ini, algoritma ID3 digunakan untuk melakukan perhitungan terhadap atribut dari hasil uji laboratorium Tumor Marker pada pasien terduga Kanker Payudara. Atribut tersebut harus dipilih yang sangat mempengaruhi terhadap penyakit Kanker Payudara. Terdapat dua cara dalam perhitungan yang harus dilakukan yaitu, perhitungan nilai Entropy terhadap masing – masing nilai atribut dari

sampel yang didapatkan. Rumus Entropy :

$$Entropy(S) \equiv \sum_i^c -p_i \log_2 p_i$$

Keterangan :

- c = jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi)
- Pi = jumlah sampe untuk kelas i.

Serta dilakukan perhitungan terhadap nilai Information Gain dari setiap atribut terkait setelah nilai Entropy diperoleh untuk dapat membentuk sebuah Pohon Keputusan yang tepat. rumus Information Gain :

$$Gain(S,A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{S} En$$

Keterangan dari rumus tersebut :

- A = atribut
- V = menyatakan nilai yang mungkin untuk atribut A.
- Values (A) = himpunan nilai – nilai yang mungkin untuk atribut A.
- |Sv| = jumlah sampel untuk nilai v.
- |S| = jumlah nilai untuk seluruh sampel data.
- Entropy (Sv) = entropy untuk sampel – sampel yang memiliki nilai v.

2.2 DECISION TREE LEARNING

Decision Tree Learning merupakan suatu metode belajar yang paling sangat ultiva dan digunakan secara praktis. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur

berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari pohon keputusan atau Decision Tree, antara lain :

- a. dalam melakukan Kemampuan break – down proses pengambilan keputusan yang kompleks dan rumit menjadi lebih sederhana.
- b. Berfungsi untuk mengeksplorasi data.
- c. Menemukan hubungan tersembunyi antara calon variabel input dan calon variabel target.
- d. Memadukan antara eksplorasi data dan eksplorasi dan pemodelan.

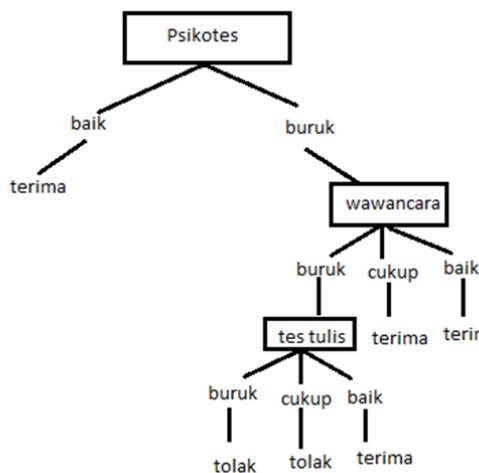
Sedangkan kelebihan yang dimiliki oleh decision tree adalah :

- a. Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih ultiv dan spesifik.
- b. Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
- c. Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama. Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
- d. Dalam analisis ultivariate, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk

mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan

Kekurangan dari Decision Tree yaitu :

- Terjadi overlap terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak. Hal tersebut juga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
 - Pengakumulasi jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.
 - Hasil kualitas keputusan yang didapatkan dari metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut didesain.
- Simulasi Decision Tree :
- Kesulitan dalam mendesain pohon keputusan yang optimal.



2.3 TUMOR MARKER

Tumor marker merupakan senyawa yang ditemukan di atas jumlah normal di dalam urin, darah, atau cairan tubuh yang lain bila terdapat kanker tertentu di dalam tubuh. Penanda Tumor atau Tumor Marker yang berasal dari sumber yang sama yaitu :

- CA-15-3 yang menunjukkan pertumbuhan sel kanker yang bermetastasis dalam tubuh manusia. Ca-15-3 biasanya lebih sensitif pada kasus kanker payudara, terutama apabila pasien sudah pernah melakukan operasi pengangkatan atau pasca-operasi. Digunakan untuk menghitung kemungkinan kekambuhan dari penyakit kanker yang ada dalam tubuh manusia. Nilai Rujukan yang digunakan dalam hasil laboratorium untuk Penanda Tumor ini adalah 0,00 - 22,00 U/mL.
- CA 27.29 merupakan pengembangan dari Ca-15-3. Penanda Tumor ini memiliki tingkat sensitifitas yang lebih tinggi dari Ca-15-3. Terutama pada kasus kanker payudara, walaupun pada kasus kanker yang lain juga tetap memiliki pengaruh atau sensitifitas tersendiri. Nilai rujukan yang digunakan dalam hasil laboratorium untuk Penanda Tumor ini adalah < 100U/mL.
- CEA (Carcinoembryonic Antigen) yang merupakan glikoprotein yang terdapat pada sel plasma darah manusia dewasa. CEA juga merupakan parameter untuk mengukur kadar zat karsinogenik dalam darah. Zat karsinogenik adalah zat yang memicu untuk tumbuhnya sel kanker dalam tubuh

manusia. Nilai Rujukan yang digunakan dalam hasil laboratorium untuk Penanda Tumor ini adalah <5 ng/mL.

Atribut Tumor Marker	Nilai Rujukan
CEA (Carcinoembyonic Antigen)	0 – 3 ng/mL (rendah) 4 – 10 ng/mL (tinggi)
Ca – 15 – 3 (Cancer Antigen 15-3)	0 – 11 U/mL (rendah) 12 – 22 U/mL (normal) 23 – 30 U/mL (tinggi)
Ca 27.29 (Cancer Antigen 27.29)	0 – 35 U/mL (rendah) 36 – 71 U/mL (normal) 72 – 110 U/mL (tinggi)

2.4 LOGIKA IF – THEN

Logika if – then digunakan dalam kasus ini, untuk menentukan tinggi rendahnya nilai kadar atribut tumor marker yang berfungsi untuk rules yang akan terbentuk pada aplikasi yang akan diciptakan.

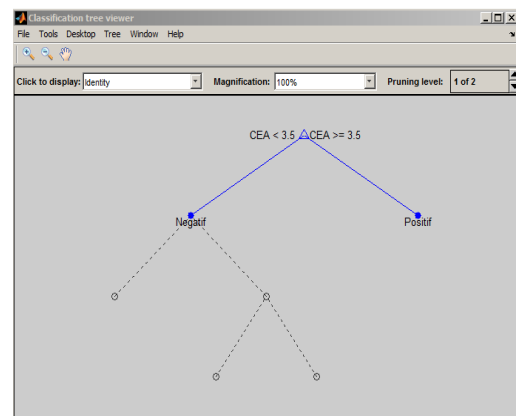
- If $CEA \leq 3 \rightarrow$ ‘Rendah’
- If $CEA > 3 \rightarrow$ ‘Tinggi’
- If $Ca - 15 - 3 \leq 11 \rightarrow$ ‘Rendah’
- If $11 < Ca - 15 - 3 \leq 22 \rightarrow$ ‘Normal’
- If $22 < Ca - 15 - 3 \leq 30 \rightarrow$ ‘Tinggi’
- If $Ca.27.29 \leq 35 \rightarrow$ ‘Rendah’
- If $35 < Ca.27.29 \leq 71 \rightarrow$ ‘Normal’
- If $71 < Ca.27.29 \leq 110 \rightarrow$ ‘Tinggi’

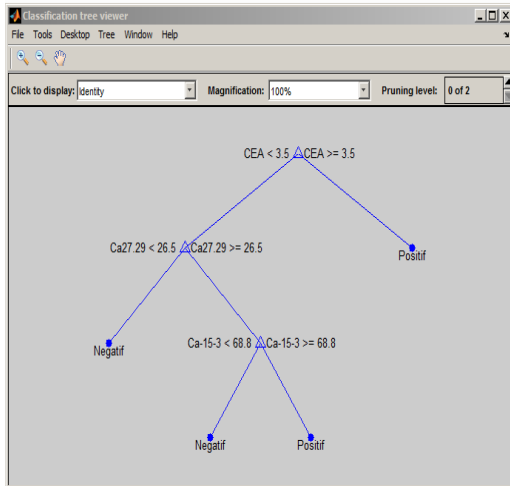
3. HASIL DAN PEMBAHASAN

Decision Tree membentuk root terhadap setiap atribut sesuai nilai Entropy dan Informtion Gain masing – masing. Dari root tersebut akan muncul dan membuktikan bahwa penelitian telah sesuai dengan implementasi yang ada.

3.1 PEMBENTUKAN ROOT

Berikut adalah hasil pembentukan root atau cabang dari setiap pohon keputusan. Root tersebut membentuk sebuah rules baru yang akan di implementasikan pada aplikasi hasil penelitian. Root yang ditunjukkan pada gambar dibawah ini, adalah root pada level 1 iterasi ke 1 dari Decision Tree yang juga menunjukkan akan terdapat dua buah cabang lagi yang akan terbentuk dari data tersebut.





Sedangkan pada gambar di atas menunjukkan keseluruhan dari Decision Tree yang terbentuk dari data random yang telah di olah dan di inputkan ke dalam Matlab.

3.2 RULES DECISION TREE

Berikut adalah rules if – then pada setiap atribut pembentuk Pohon Keputusan yang di olah menggunakan Matlab :

```
t =
Decision tree for classification
1 if CEA<3.5 then node 2 elseif CEA>=3.5 then node 3 else Positif
2 if Ca27.29<26.5 then node 4 elseif Ca27.29>=26.5 then node 5
3 class = Positif
4 class = Negatif
5 if Ca-15-3<68.8 then node 6 elseif Ca-15-3>=68.8 then node 7
6 class = Negatif
7 class = Positif
...

```

3.3 IMPLEMENTASI APLIKASI

Berikut merupakan hasil akhir dari penelitian yang dilakukan yaitu, pembentukan sebuah aplikasi menggunakan Matlab.



Dibawah ini, merupakan hasil akhir beserta keterangan pada setiap bagian dan tools nya, serta menunjukkan cara kerja aplikasi tersebut.



4. KESIMPULAN DAN SARAN

- 1) Algoritma ID3 yang digabungkan dengan pembentukan Decision Tree dapat digunakan sebagai pendukung keputusan dari diagnosa penyakit Kanker Payudara. Perhitungan menggunakan Algoritma ID3 hingga terbentuknya Decision Tree terhadap data random yang digunakan saat dilakukan ujicoba menghasilkan tingkat akurasi mencapai 100%.
- 2) Penelitian selanjutnya dapat menggunakan Algoritma

- Klasifikasi Data Mining yang lain selain dari Algoritma ID3.
- 3) Dapat melakukan Deteksi dan Analisa terhadap penyakit lain baik kanker maupun bukan dengan menggunakan Algoritma ID3 dan atau Algoritma Klasifikasi Data Mining lainnya.

DAFTAR PUSTAKA

1. *Penanda tumor-Wikipedia bahasa Indonesia*. (2014, Maret 12). Retrieved Maret 12, 2015, from Wikipedia Indonesia: http://id.m.wikipedia.org/wiki/Penanda_tumor
2. Defiyanti, Sofi; Pardede, D. L. Crispina;. (n.d.). PERBANDINGAN KINERJA ALGORITMA ID3 DAN C4.5. 1-5.
3. F.Romansyah; Sitanggang, I.S; S.Nurdiati;. (2009). Fuzzy Decision Tree dengan Algoritma ID3 pada data Diabetes. *Internetworking Indonesia Journal*, 1-8.
4. JThomas;. (2015, Maret 08). *Kanker-Wikipedia bahasa Indonesia*. Retrieved Maret 13, 2015, from Wikipedia Indonesia: <http://id.m.wikipedia.org/wiki/Kanker>
5. Maharani, Sabrina;. (2009). *Mengenal 13 jenis Kanker & pengobatannya*. Jogjakarta: KATAHATI.
6. Monash University, Faculty of Information Technology;. (2004). CSE5230 Tutorial: The ID3 Decision Tree Algorithm. *CSE5230 Data Mining*, 1-8.
7. Peng, Wei; Chen, Juhua; Zhou, Haiping;. (2010). An Implementation of ID3--- Decision Tree Learning Algorithm. *Project of Comp 9417 : Machine Learning*, 3-16.
8. PHK TIK K1, Universitas Widyagama;. (2008). Logika Fuzzy. 3-26.
9. Presetyo, Eko;. (2014). *Mengolah Data Menjadi Informasi Menggunakan Matlab*. Jogjakarta: ANDI.
10. Primartha, Rifkie; Fathiyah, Nurul;. (2012). SISTEM PAKAR FUZZY UNTUK DIAGNOSIS KANKER PAYUDARA. *Jurnal Generic, Vol. x, No. x, Bulan 201x, pp. x~x*, 1-14.
11. (n.d.). programming in Java. In *ID3 : Learning from Examples* (pp. 369-388).
12. S.Si,M.Kom, T.Sutojo; S.Si,M.Kom, Edy Mulyanto; Suhartono, Dr. Vincent;. (2011). *Kecerdasan Buatan*. Jogjakarta: ANDI.
13. Santosa, Budi;. (2007). *Data Mining Terapan dengan Matlab*. Jogjakarta: GRAHA ILMU.
14. Smart, Aqila;. (2010). *Kanker Organ Reproduksi*. Jogjakarta: A'PLUS BOOKS.
15. Stein, & Letmann. (2005-2013). Chapter ML : III. In Stein, & Letmann, *Decision Trees* (pp. 64-88).
16. Wahyudin;. (2009). Metode Iterative Dichotomizer 3 (ID3)

Untuk Penyeleksian Penerimaan
Mahasiswa Baru. *Jurnal*

*Pendidikan Teknologi Informasi
dan Komunikasi, 5-15.*