

## Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis

Jefry Latu Handarko<sup>✉</sup>, Alamsyah

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia  
Gedung D7 Lt.1, Kampus Sekaran Gunungpati, Semarang 50299

### Info Artikel

*Sejarah Artikel:*

Diterima Januari 2015

Disetujui Januari 2015

Dipublikasikan November 2015

*Keywords:*

*Fuzzy;*

*Decision Tree;*

*Algoritma ID3;*

*MySQL..*

### Abstrak

Tujuan yang ingin dicapai dalam penelitian ini adalah (1) Untuk menerapkan salah satu teknik klasifikasi yaitu *Fuzzy ID3 Decision Tree* pada data pemeriksaan hasil lab pasien; (2) Untuk mengetahui hasil implementasi dari *Fuzzy ID3* pada data hasil lab pasien menggunakan aplikasi *PHP MySQL* yang sudah dirancang. Metode yang digunakan dalam penelitian ini yaitu metode klasifikasi dengan *fuzzy decision tree*. Dengan menerapkan teknik *data mining* pada data hepatitis diharapkan dapat ditemukan aturan klasifikasi yang dapat digunakan untuk memprediksi potensi seseorang terserang penyakit hepatitis. Algoritma yang digunakan dalam *fuzzy decision tree* adalah ID3. Hasil Implementasi *Fuzzy ID3* terhadap data hepatitis B adalah sebagai berikut: (a) Ditentukannya aturan fuzzy untuk ketiga training set; (b) perhitungan ketiga training set diperoleh keakuratan yang terbaik yaitu dengan nilai 88,5% dimana data yang dipakai 15 data *training set*; (c) Pembentukan *Fuzzy ID3* sangat berpengaruh pada hasil *training set*. Semakin banyak dan semakin akurat datanya maka akan menaikkan keakuratan hasil *Fuzzy ID3*nya.

### Abstract

The objectives of this research are (1) To apply one of the classification techniques, namely *Fuzzy ID3 Decision Tree* on the data examination patient lab results; (2) To know the results of the implementation of *Fuzzy ID3* lab results patient data using *MySQL PHP* application that has been designed. The method used in this research is the method of classification by *fuzzy decision tree*. By applying *data mining* techniques on the data expected to be found hepatitis classification rules that can be used to predict a person's potential disease hepatitis. The algorithm used in the *fuzzy decision tree* is ID3. Results Implementation of *Fuzzy ID3* against hepatitis B the data are as follows: (a) it determines the fuzzy rules for the third training set; (b) calculation of the training set third best accuracy is obtained with a value of 88.5% where data used 15 training data sets; (c) Establishment of *Fuzzy ID3* influence on the outcome of the training set. The more and more accurate data will increase the accuracy of the results of *Fuzzy ID3*.

## Pendahuluan

Dengan adanya kemajuan teknologi saat ini, suatu penyakit akan terdeteksi dengan lebih cepat melalui gejala-gejala tersebut. Tidak hanya hal tersebut, dalam hal proses diagnosa dikenal dengan adanya uji tes darah di laboratorium untuk mengetahui jenis penyakit yang diderita oleh pasien secara pasti.

Hepatitis adalah suatu penyakit peradangan hati yang umumnya disebabkan oleh virus, yang dewasa ini banyak diderita baik orang dewasa mau pun anak-anak (Rosalina, 2012). Karena itu, diperlukan suatu penanganan dini sebelum penderita penyakit tersebut bertambah parah. Salah satu cara penanganan dini dari penyakit hepatitis ini adalah melakukan pemeriksaan sejak awal, karena itu dibuatlah suatu sistem diagnosa penyakit hepatitis yang dapat membantu penderita untuk melakukan pemeriksaan.

Klasifikasi merupakan salah satu metode dalam *data mining* untuk memprediksi label kelas dari suatu *record* dalam data (Ridwan dkk, 2013). Klasifikasi hepatitis sebenarnya dapat dilakukan dengan menggunakan logika tegas. Akan tetapi hal ini sangat kaku, karena dengan adanya perubahan yang kecil saja terhadap nilai dapat mengakibatkan perbedaan kategori. Logika *fuzzy* digunakan untuk mengantisipasi hal tersebut, karena dapat memberikan toleransi terhadap nilai, sehingga dengan adanya perubahan sedikit pada nilai tidak akan memberikan perbedaan yang signifikan. Dengan memanfaatkan kelebihan logika *fuzzy* dalam toleransi terhadap hal ambigu, diharapkan dapat menjadi pendukung keputusan dalam mengklasifikasikan data hepatitis.

*Fuzzy* secara bahasa diartikan sebagai kabur atau samar samar. Suatu nilai dapat bernilai benar atau salah secara bersamaan. Dalam *fuzzy* dikenal derajat keanggotaan yang memiliki rentang nilai 0 (nol) hingga 1 (satu). Berbeda dengan himpunan tegas yang memiliki nilai 1 atau 0 (ya atau tidak). *Logika Fuzzy* merupakan suatu *logika* yang memiliki nilai kekaburan atau kesamaran (*fuzzyness*) antara benar atau salah. Dalam teori *logika fuzzy* suatu nilai bisa bernilai benar atau salah secara bersamaan. Namun seberapa besar keberadaan dan kesalahannya tergantung pada bobot keanggotaan yang dimilikinya (Kusumadewi & Purnomo, 2003).

Himpunan adalah suatu kumpulan atau koleksi objek-objek yang mempunyai kesamaan sifat tertentu. Himpunan *fuzzy* merupakan suatu pengembangan lebih lanjut tentang konsep

himpunan dalam matematika. Himpunan *fuzzy* adalah rentang nilai-nilai, masing-masing nilai mempunyai derajat keanggotaan antara 0 sampai dengan 1. Prinsip dasar dan persamaan matematika dari teori himpunan fuzzy adalah sebuah teori pengelompokan objek dalam batas yang samar. Himpunan tersebut dikaitkan dengan suatu fungsi yang menyatakan derajat kesesuaian unsur-unsur dalam semestanya dengan konsep yang merupakan syarat keanggotaan himpunan tersebut. Fungsi itu disebut *fungsi keanggotaan* dan nilai fungsi itu disebut derajat keanggotaan suatu unsur dalam himpunan itu, yang selanjutnya disebut himpunan kabur (*fuzzy set*) (Susilo, 2003).

Ada beberapa fungsi yang bisa digunakan (a) Representasi Linear, Pada representasi linear, pemetaan input ke derajat keanggotaannya dapat digambarkan sebagai suatu garis lurus. Bentuk ini paling sederhana dan menjadi pilihan yang baik untuk mendekati suatu konsep yang kurang jelas. Ada 2 keadaan himpunan *fuzzy* yang linear. Pertama, Representasi linear naik, yaitu kenaikan himpunan dimulai dari nilai domain yang memiliki nilai keanggotaan nol [0] bergerak ke kanan menuju ke nilai domain yang memiliki derajat keanggotaan yang lebih tinggi. Kedua, representasi linear turun yaitu garis lurus yang dimulai dari nilai domain dengan derajat keanggotaan tertinggi pada sisi kiri, kemudian bergerak turun ke nilai domain yang memiliki derajat keanggotaan lebih rendah. (b) Representasi Kurva Segitiga Representasi kurva segitiga, pada dasarnya adalah gabungan antara dua representasi linear (representasi linear naik dan representasi linear turun). (c) Representasi Kurva Trapesium Kurva Trapesium pada dasarnya seperti bentuk segitiga, hanya saja ada beberapa titik yang memiliki nilai keanggotaan 1.

*Decision tree* merupakan suatu pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan permasalahan *klasifikasi*. *Data mining* merupakan solusi yang mampu menemukan kandungan informasi yang tersembunyi berupa pola dan aturan dari sekumpulan data yang besar agar mudah dipahami. Informasi yang tersembunyi ini sangat menguntungkan dari sudut pandang penelitian, bisnis dan lainnya. *Decision tree* merupakan suatu pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan permasalahan *klasifikasi*. (Romansyah, 2009).

*Information gain* adalah suatu nilai statistik yang digunakan untuk memilih atribut yang akan mengekspansi *tree* dan menghasilkan *node* baru pada algoritma FID3. Suatu *entropy* dipergunakan untuk mendefinisikan nilai *information gain*.

*Entropy* dirumuskan sebagai berikut:

$$H_s(s) = \sum_i^N -p_i * \log_2(P_i)$$

Dengan  $P_i$  adalah *rasio* dari kelas  $C_i$  pada himpunan. Contoh  $S = \{x_1, x_2, \dots, x_k\}$ .

*Information gain* digunakan sebagai ukuran seleksi atribut, yang merupakan hasil pengurangan *entropy* dari himpunan.

*Informasi gain* didefinisikan sebagai berikut :

$$G_f(S) = H_f(S) - \sum_{v \in A} \frac{|S_v|}{|S|} * H_f(S_v, A)$$

Dengan  $\mu_j$  adalah nilai keanggotaan dari pola ke- $j$  untuk kelas ke- $i$ .  $H_j(S)$  menunjukkan *entropy* dari himpunan  $S$  dari data pelatihan pada *node*.  $|S_v|$  adalah ukuran dari subset  $S_v \subseteq S$  dari pelatihan  $x_j$  dengan atribut  $v$ .  $|S|$  menunjukkan ukuran dari himpunan  $S$ .

Jika pada proses *learning* dari FDT dihentikan sampai semua data contoh pada masing-masing *leaf-node* menjadi anggota sebuah kelas, akan dihasilkan akurasi yang rendah. Oleh karena itu untuk meningkatkan akurasinya, proses *learning* harus dihentikan lebih awal atau melakukan pemotongan *tree* secara umum (Romansyah, 2009).

Jika pada proses *learning* dari FDT dihentikan sampai semua data contoh pada masing-masing *leaf-node* menjadi anggota sebuah kelas, akan dihasilkan akurasi yang rendah. Oleh karena itu untuk meningkatkan akurasinya, proses *learning* harus dihentikan lebih awal atau melakukan pemotongan *tree* secara umum (Romansyah, 2009). Untuk itu diberikan 2 (dua) *threshold* yang harus terpenuhi jika *tree* akan diekspansi, yaitu: (a) *Fuzziness control threshold* (FCT) /  $\theta_r$ , Jika proporsi dari himpunan data dari kelas  $C_k$  lebih besar atau sama dengan nilai *threshold*  $\theta_r$ , maka hentikan ekspansi *tree*. Sebagai contoh: jika pada sebuah *sub-data set* rasio dari kelas 1 adalah 90%, kelas 2 adalah 10% dan  $\theta_r$  adalah 85%, maka hentikan ekspansi *tree*. (b) *Leaf decision threshold* (LDT) /  $\theta_n$  Jika banyaknya anggota himpunan data pada suatu *node* lebih kecil dari *threshold*  $\theta_n$ , hentikan

ekspansi *tree*. Sebagai contoh: sebuah himpunan data memiliki 600 contoh dengan  $\theta_n$  adalah 2%. Jika jumlah data contoh pada sebuah *node* lebih kecil dari 12 (2% dari 600), maka hentikan ekspansi *tree*.

Program yang digunakan adalah *PHP MySQL*. *MySQL* juga dapat berjalan pada personal komputer (banyak pengembangan dari *MySQL* terjadi pada system yang tidak mahal yaitu *Linux System*). Tetapi *MySQL* juga *portable* dan dapat berjalan pada system operasi yang komersial seperti misalnya *Windows, Solaris, Irix*. *MySQL* menggunakan bahasa *SQL*.

### Metode

Metode yang digunakan dalam penelitian ini yaitu metode klasifikasi dengan *fuzzy decision tree*. Penggunaan teknik *fuzzy* memungkinkan melakukan prediksi suatu objek yang dimiliki oleh lebih dari satu kelas. Dengan menerapkan teknik *data mining* pada data hepatitis diharapkan dapat ditemukan aturan klasifikasi yang dapat digunakan untuk memprediksi potensi seseorang terserang penyakit hepatitis.

*Decision tree* membuat aturan *rule* yang dapat digunakan untuk menentukan apakah seseorang mempunyai potensi untuk menderita Hepatitis atau tidak. Algoritma yang digunakan dalam *fuzzy decision tree* adalah ID3 (*Iterative Dichotomiser 3*). Algoritma ini menggunakan teori informasi untuk menentukan atribut mana yang paling informatif. Namun ID3 (*Iterative Dichotomiser 3*) sangat tidak stabil dalam melakukan klasifikasi yang berkenaan dengan gangguan kecil pada data pelatihan. Logika *fuzzy* dapat memberikan suatu peningkatan dalam melakukan klasifikasi pada saat pelatihan.

Metode ini digunakan untuk memperkirakan nilai diskret dari fungsi target, yang mana fungsi pembelajaran direpresentasikan oleh sebuah *decision tree*. *Decision tree* merupakan himpunan aturan IF...THEN. Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, di mana premis terdiri atas sekumpulan *node-node* yang ditemui, dan kesimpulan dari aturan terdiri atas kelas yang terhubung dengan *leaf* dari *path*

*Fuzzy decision tree* memungkinkan untuk menggunakan nilai-nilai *numeric-symbolic* selama konstruksi atau saat mengklasifikasikan kasus-kasus baru. Manfaat dari teori himpunan *fuzzy* dalam *decision tree* ialah meningkatkan kemampuan dalam memahami *decision tree* ketika menggunakan atribut-atribut kuantitatif. Bahkan, dengan menggunakan teknik *fuzzy* dapat

meningkatkan ketahanan saat melakukan klasifikasi kasus-kasus baru.

FID3 (*Fuzzy Iterative Dichotomiser 3*) merupakan salah satu algoritma yang banyak digunakan untuk membuat suatu *decision tree*. Algoritma ini pertama kali diperkenalkan oleh Quinlan, menggunakan teori informasi untuk menentukan atribut mana yang paling informatif, namun FID3 sangat tidak stabil dalam melakukan penggolongan berkenaan dengan gangguan kecil pada data pelatihan. Logika *fuzzy* dapat memberikan suatu peningkatan untuk dalam melakukan penggolongan pada saat pelatihan (Romansyah, 2009:47).

**Pembahasan**

Pada tahap ini dilakukan teknik *data mining* menggunakan algoritma ID3 untuk membangun *fuzzy decision tree* (FDT). Hal pertama yang dilakukan adalah membentuk training set. Setelah training set dibentuk, maka pembentukan *decision tree* bisa dilakukan dengan melakukan memilih *root node* dan *node-node* lainnya menggunakan perhitungan *fuzzy entropy* dan *information gain*.

Data yang telah ditransformasi akan dibuat menjadi data training (training set). Training set yang pertama dibentuk menggunakan 10 data seperti pada Tabel 1.

Tabel 1. 10 Data *Training Set*

NO	pasien	umur	HbeAg	HBV DNA	Alt	diagnosa
1	p1	15	positif	rendah	tinggi	tidak
2	p2	20	positif	sedang	sedang	ya
3	p3	20	positif	tinggi	sedang	ya
4	p4	40	positif	sedang	rendah	tidak
5	p5	45	positif	tinggi	sedang	ya
6	p6	17	negatif	rendah	tinggi	tidak
7	p7	50	negatif	tinggi	sedang	ya
8	p8	70	negatif	sedang	tinggi	ya
9	p9	80	negatif	sedang	tinggi	ya
10	p10	75	negatif	tinggi	rendah	ya

Training set berikutnya dibentuk menggunakan 15 data yang tampak seperti pada Tabel 2 dan Training set menggunakan 20 data yang tampak seperti pada Tabel 3.

Tabel 2. 15 Data *Training Set*

NO	pasien	umur	HbeAG	HBV DNA	Alt	diagnosa
1	p1	15	negatif	tinggi	sedang	ya
2	p2	30	negatif	sedang	rendah	ya
3	p3	45	positif	tinggi	rendah	ya
4	p4	50	positif	rendah	rendah	tidak
5	p5	12	negatif	tinggi	rendah	ya
6	p6	35	negatif	sedang	sedang	ya
7	p7	80	positif	sedang	rendah	ya
8	p8	16	positif	rendah	rendah	tidak
9	p9	60	negatif	tinggi	tinggi	ya
10	p10	75	negatif	tinggi	rendah	ya
11	p11	40	negatif	rendah	rendah	tidak
12	p12	50	positif	sedang	tinggi	ya
13	p13	19	positif	sedang	tinggi	tidak
14	p14	45	positif	tinggi	tinggi	ya
15	p15	30	negatif	tinggi	rendah	ya

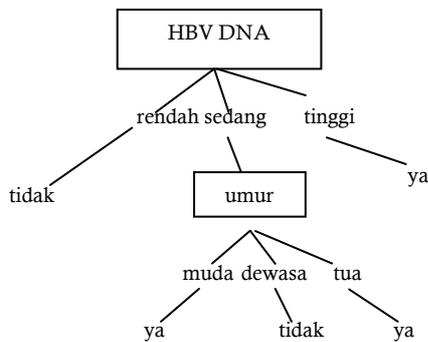
Tabel 3. 20 Data *Training Set*

NO	pasien	umur	HbeAG	HBV DNA	Alt	diagnosa
1	p1	10	positif	rendah	rendah	tidak
2	p2	20	positif	rendah	sedang	tidak
3	p3	25	negatif	rendah	sedang	tidak
4	p4	30	negatif	sedang	tinggi	tidak
5	p5	25	positif	rendah	rendah	tidak
6	p6	15	negatif	tinggi	tinggi	ya
7	p7	76	positif	tinggi	sedang	ya
8	p8	40	negatif	tinggi	sedang	ya
9	p9	50	positif	tinggi	tinggi	ya
10	p10	45	positif	sedang	tinggi	tidak
11	p11	35	positif	tinggi	sedang	ya
12	p12	80	negatif	tinggi	tinggi	ya
13	p13	35	negatif	sedang	tinggi	tidak
14	p14	37	positif	tinggi	tinggi	ya
15	p15	30	negatif	tinggi	sedang	ya
16	p16	28	positif	rendah	sedang	tidak
17	p17	15	negatif	tinggi	sedang	ya
18	p18	31	positif	sedang	tinggi	tidak
19	p19	55	negatif	tinggi	tinggi	ya
20	p20	45	positif	sedang	sedang	ya

Setelah dibentuk training setnya, mulailah dibentuk *root node*nya. Untuk membentuk *root*

nodanya bisa dilakukan dengan memilih hasil *Information gain* terbesarnya. Pada 10 data training set didapat, *information gain* variabel umur, *HbeAG*, *HBV DNA* dan *Alt* masing-masing adalah 0.2473, 0.0349, 0.5568 dan 0.2813. Dari hasil tersebut maka *root node* dari 10 data *training set* adalah *HBV DNA* karena memiliki *information gain* terbesar.

Lakukan terus ekspansi dari *sub-node* yang ada sampai tidak ada lagi data yang dapat diekspansi atau tidak ada lagi atribut yang dapat digunakan untuk mengekspansi *tree* yaitu ketika *tree* yang terbentuk sudah mencapai kedalaman maksimum atau *sub-node* tidak memenuhi syarat dari *threshold* yang diberikan. Jika *sub-node* sudah tidak dapat diekspansi maka nilai proporsi kelas terbesar merupakan kesimpulan dari sekumpulan aturan yang diperoleh dengan menghubungkan setiap *node* yang dilewati dari *root node* hingga *leaf node*. Gambar 1 menunjukkan *decision tree* untuk 10 data *training set*.



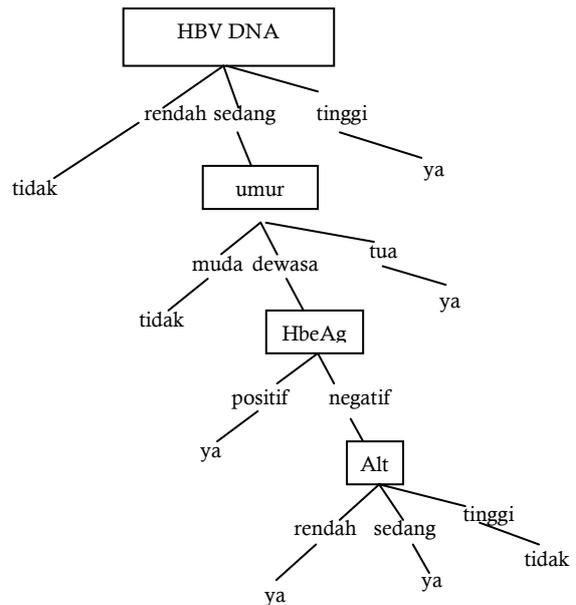
Gambar 1. Pohon Keputusan 10 Data Training Set

Pada 15 data training set didapat, *information gain* variabel umur, *HbeAG*, *HBV DNA* dan *Alt* masing-masing adalah 0.1007, 0.0258, 0.5946 dan 0.0437. Dari hasil tersebut maka *root node* dari 15 data *training set* adalah *HBV DNA* karena memiliki *information gain* terbesar. Jika *sub-node* sudah tidak dapat diekspansi maka nilai proporsi kelas terbesar merupakan kesimpulan dari sekumpulan aturan yang diperoleh dengan menghubungkan setiap *node* yang dilewati dari *root node* hingga *leaf node*.

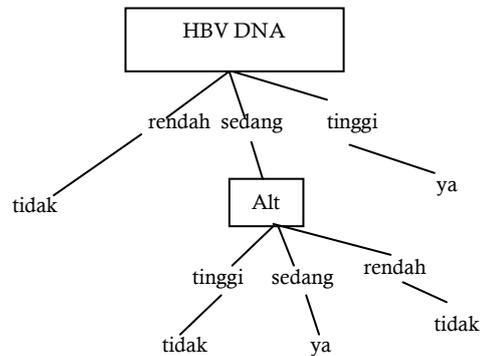
Pada 20 data training set didapat, *information gain* variabel umur, *HbeAG*, *HBV DNA* dan *Alt* masing-masing adalah 0.0928, 0.0329, 0.8123 dan 0.1336. Dari hasil tersebut maka *root node* dari 20 data *training set* adalah *HBV DNA* karena memiliki *information gain* terbesar.

Lakukan terus ekspansi dari *sub-node* yang ada sampai tidak ada lagi data yang dapat

diekspansi atau tidak ada lagi atribut yang dapat digunakan untuk mengekspansi *tree* yaitu ketika *tree* yang terbentuk sudah mencapai kedalaman maksimum atau *sub-node* tidak memenuhi syarat dari *threshold* yang diberikan. Jika *sub-node* sudah tidak dapat diekspansi maka nilai proporsi kelas terbesar merupakan kesimpulan dari sekumpulan aturan yang diperoleh dengan menghubungkan setiap *node* yang dilewati dari *root node* hingga *leaf node*.



Gambar 2. Pohon Keputusan 15 Data training set



Gambar 3. Pohon Keputusan 20 Data Training Set

Gambar 2 dan 3 merupakan hasil *fuzzy decision tree* untuk 15 dan 20 data *training set*. Pada Pengujian yang dilakukan ternyata perhitungan ID3 menggunakan 15 data training set mencapai keakurasian 100%, kemudian 10 data training set 60% sedangkan 20 data training set hanya 40%.

Peneliti kemudian meneliti dalam skala yang lebih besar, kali ini peneliti menghitung menggunakan 70 data (terlampir) dan hasilnya

adalah pada data 10 training set terdapat 57 data benar sehingga keakurasiannya 81.4%. Kemudian pada data 15 training set, terdapat 62 data benar sehingga keakurasiannya 88.5%. Dan pada 20 data training set yang digunakan hanya 54 data benar dan mencapai keakurasiannya 77.1%.

Jadi perhitungan paling akurat menggunakan perhitungan ID3 dengan 15 data training set. Pembentukan training set ini diambil data yang berjumlah 15 dan terdapat variasi didalamnya misalnya pada variabel Alt diambil tidak hanya kebanyakan rendah atau tinggi tetapi terdapat rendah, sedang dan tinggi dengan jumlah yang hampir sama.

Pada masalah ini bisa dilihat perhitungan ID3 sangat dipengaruhi oleh pembentukan training set itu sendiri. Pada training set pertama dibentuk hanya menggunakan 10 data saja tetapi bervariasi isinya. Pada training set kedua dibentuk menggunakan 15 data dan bervariasi. Dan training set ketiga dibentuk menggunakan 20 data tetapi sangat monoton.

Tingkat keakurasiannya perhitungan ID3 ini dipengaruhi oleh training set yang dibuat. Semakin training set yang bervariasi datanya dan semakin banyak pula datanya maka tingkat akurasi perhitungan ID3 ini juga akan semakin akurat. Tampilan aplikasi untuk perhitungan *fuzzy entropy* dan *information gain* pada 10 data training set seperti pada Gambar 4.

The screenshot shows a web browser window with the URL `localhost:8080/cekhepatitis/cekhepatitis/RumusDasar.php?act=entropi&aksi=10hasil`. The page displays two data tables and a calculation summary.

**Tabel 10 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	20	Positif	Rendah	Tinggi	Tidak
2	p2	30	Positif	Sedang	Sedang	Ya
3	p3	30	Positif	Tinggi	Sedang	Ya
4	p4	40	Positif	Sedang	Rendah	Tidak
5	p5	45	Positif	Tinggi	Sedang	Ya
6	p6	20	Negatif	Rendah	Tinggi	Tidak
7	p7	50	Negatif	Tinggi	Sedang	Ya
8	p8	60	Negatif	Sedang	Tinggi	Ya
9	p9	60	Negatif	Sedang	Tinggi	Ya
10	p10	75	Negatif	Tinggi	Rendah	Ya

**Tabel 15 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	15	Negatif	Tinggi	Sedang	Ya
2	p2	30	Negatif	Sedang	Rendah	Ya
3	p3	45	Positif	Tinggi	Rendah	Ya
4	p4	50	Positif	Rendah	Rendah	Tidak
5	p5	12	Negatif	Tinggi	Rendah	Ya
6	p6	35	Negatif	Sedang	Tinggi	Tidak
7	p7	80	Positif	Sedang	Rendah	Ya
8	p8	15	Positif	Rendah	Rendah	Tidak

**Hasil Entropi & Gain Tabel 10 Data**

Diagnosa: sampel kelas 1 'Ya' = 7, sampel kelas 2 'Tidak' = 3

Umur: Umur Muda 'Ya' = 2 -> 'Tidak' = 2, Umur Dewasa 'Ya' = 2 -> 'Tidak' = 1, Umur Tua 'Ya' = 3 -> 'Tidak' = 0

HbeAG: HbeAG Positif 'Ya' = 3 -> 'Tidak' = 2, HbeAG Negatif 'Ya' = 4 -> 'Tidak' = 1

HBV-DNA: HBV-DNA Rendah 'Ya' = 0 -> 'Tidak' = 3, HBV-DNA Sedang 'Ya' = 3 -> 'Tidak' = 1, HBV-DNA Tinggi 'Ya' = 4 -> 'Tidak' = 0

Alt: Alt Rendah 'Ya' = 1 -> 'Tidak' = 1, Alt Sedang 'Ya' = 4 -> 'Tidak' = 0, Alt Tinggi 'Ya' = 2 -> 'Tidak' = 2

Perhitungan Entropi & Gain

Entropi Total: 0.8813

Umur: Entropi Umur Muda: 1, Entropi Umur Dewasa: 0.9183, Entropi Umur Tua: 0, Gain Umur: 0.2058

HbeAG: Entropi HbeAG Positif: 0.971, Entropi HbeAG Negatif: 0.7219, Gain HbeAG: 0.0349

HBV-DNA: Entropi HBV-DNA Rendah: 0, Entropi HBV-DNA Sedang: 0.8113

Gambar 4. Tampilan Hasil 10 Data Training Set

Kemudian untuk perhitungan *fuzzy entropy* dan *information gain* pada 15 data training set tampak seperti Gambar 5.

The screenshot shows a web browser window with the URL `localhost:8080/cekhepatitis/cekhepatitis/RumusDasar.php?act=entropi&aksi=15hasil`. The page displays two data tables and a calculation summary.

**Tabel 10 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	20	Positif	Rendah	Tinggi	Tidak
2	p2	30	Positif	Sedang	Sedang	Ya
3	p3	30	Positif	Tinggi	Sedang	Ya
4	p4	40	Positif	Sedang	Rendah	Tidak
5	p5	45	Positif	Tinggi	Sedang	Ya
6	p6	20	Negatif	Rendah	Tinggi	Tidak
7	p7	50	Negatif	Tinggi	Sedang	Ya
8	p8	60	Negatif	Sedang	Tinggi	Ya
9	p9	60	Negatif	Sedang	Tinggi	Ya
10	p10	75	Negatif	Tinggi	Rendah	Ya

**Tabel 15 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	15	Negatif	Tinggi	Sedang	Ya
2	p2	30	Negatif	Sedang	Rendah	Ya
3	p3	45	Positif	Tinggi	Rendah	Ya
4	p4	50	Positif	Rendah	Rendah	Tidak
5	p5	12	Negatif	Tinggi	Rendah	Ya
6	p6	35	Negatif	Sedang	Tinggi	Tidak
7	p7	80	Positif	Sedang	Rendah	Ya
8	p8	15	Positif	Rendah	Rendah	Tidak

**Hasil Entropi & Gain Tabel 15 Data**

Diagnosa: sampel kelas 1 'Ya' = 10, sampel kelas 2 'Tidak' = 5

Umur: Umur Muda 'Ya' = 2 -> 'Tidak' = 2, Umur Dewasa 'Ya' = 6 -> 'Tidak' = 3, Umur Tua 'Ya' = 2 -> 'Tidak' = 0

HbeAG: HbeAG Positif 'Ya' = 4 -> 'Tidak' = 3, HbeAG Negatif 'Ya' = 6 -> 'Tidak' = 2

HBV-DNA: HBV-DNA Rendah 'Ya' = 0 -> 'Tidak' = 3, HBV-DNA Sedang 'Ya' = 3 -> 'Tidak' = 2, HBV-DNA Tinggi 'Ya' = 7 -> 'Tidak' = 0

Alt: Alt Rendah 'Ya' = 0 -> 'Tidak' = 3, Alt Sedang 'Ya' = 1 -> 'Tidak' = 3, Alt Tinggi 'Ya' = 3 -> 'Tidak' = 2

Perhitungan Entropi & Gain

Entropi Total: 0.9183

Umur: Entropi Umur Muda: 1, Entropi Umur Dewasa: 0.9183, Entropi Umur Tua: 0, Gain Umur: 0.1007

HbeAG: Entropi HbeAG Positif: 0.8113, Entropi HbeAG Negatif: 0.9852, Gain HbeAG: 0.0258

HBV-DNA: Entropi HBV-DNA Rendah: 0, Entropi HBV-DNA Sedang: 0.971, Entropi HBV-DNA Tinggi: 0

Gambar 5. Tampilan Hasil 15 Data Training Set

Lalu untuk perhitungan *fuzzy entropy* dan *information gain* pada 20 data training set tampak seperti Gambar 6.

The screenshot shows a web browser window with the URL `localhost:8080/cekhepatitis/cekhepatitis/RumusDasar.php?act=entropi&aksi=20hasil`. The page displays two data tables and a calculation summary.

**Tabel 10 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	20	Positif	Rendah	Tinggi	Tidak
2	p2	30	Positif	Sedang	Sedang	Ya
3	p3	30	Positif	Tinggi	Sedang	Ya
4	p4	40	Positif	Sedang	Rendah	Tidak
5	p5	45	Positif	Tinggi	Sedang	Ya
6	p6	20	Negatif	Rendah	Tinggi	Tidak
7	p7	50	Negatif	Tinggi	Sedang	Ya
8	p8	60	Negatif	Sedang	Tinggi	Ya
9	p9	60	Negatif	Sedang	Tinggi	Ya
10	p10	75	Negatif	Tinggi	Rendah	Ya

**Tabel 15 Data**

No	Pasien	Umur	HbeAG	HBV DNA	Alt	Diagnosa
1	p1	15	Negatif	Tinggi	Sedang	Ya
2	p2	30	Negatif	Sedang	Rendah	Ya
3	p3	45	Positif	Tinggi	Rendah	Ya
4	p4	50	Positif	Rendah	Rendah	Tidak
5	p5	12	Negatif	Tinggi	Rendah	Ya
6	p6	35	Negatif	Sedang	Tinggi	Tidak
7	p7	80	Positif	Sedang	Rendah	Ya

**Hasil Entropi & Gain Tabel 20 Data**

Diagnosa: sampel kelas 1 'Ya' = 12, sampel kelas 2 'Tidak' = 8

Umur: Umur Muda 'Ya' = 2 -> 'Tidak' = 2, Umur Dewasa 'Ya' = 8 -> 'Tidak' = 6, Umur Tua 'Ya' = 2 -> 'Tidak' = 0

HbeAG: HbeAG Positif 'Ya' = 6 -> 'Tidak' = 3, HbeAG Negatif 'Ya' = 6 -> 'Tidak' = 5

HBV-DNA: HBV-DNA Rendah 'Ya' = 0 -> 'Tidak' = 5, HBV-DNA Sedang 'Ya' = 2 -> 'Tidak' = 3, HBV-DNA Tinggi 'Ya' = 10 -> 'Tidak' = 0

Alt: Alt Rendah 'Ya' = 0 -> 'Tidak' = 2, Alt Sedang 'Ya' = 6 -> 'Tidak' = 3, Alt Tinggi 'Ya' = 6 -> 'Tidak' = 3

Perhitungan Entropi & Gain

Entropi Total: 0.971

Umur: Entropi Umur Muda: 1, Entropi Umur Dewasa: 0.9852, Entropi Umur Tua: 0, Gain Umur: 0.0814

HbeAG: Entropi HbeAG Positif: 0.9183, Entropi HbeAG Negatif: 0.994, Gain HbeAG: 0.0111

HBV-DNA: Entropi HBV-DNA Rendah: 0, Entropi HBV-DNA Sedang: 0.971, Entropi HBV-DNA Tinggi: 0

Gambar 6. Tampilan Hasil 20 Data Training Set

Kemudian untuk pengujian pada data pasien tampak seperti Gambar 7.

The screenshot shows a web browser window with the URL `localhost:8080/cekhepatitis/cekhepatitis/proses/data/pasien.php?act=pasien&aksi=hasil&n=6`. The page displays patient data and diagnosis results.

**Data Pasien**

No	Nama Pasien	Alamat	TTL	Umur	No Telp	Hbe AG	HBV DNA	Alt	Diagnosa
6	Agas	semarang tengah/semarang/20	0999	2	5	0	0	0	Pasien Terdiagnosa Hepatitis B

**Hasil Diagnosa Dengan Tabel 10 Data**

No	Nama Pasien	Alamat	TTL	Umur	No Telp	Hbe AG	HBV DNA	Alt	Diagnosa
6	Agas	semarang tengah/semarang/20	0999	2	5	0	0	0	Pasien Terdiagnosa Hepatitis B

**Hasil Diagnosa Dengan Tabel 15 Data**

No	Nama Pasien	Alamat	TTL	Umur	No Telp	Hbe AG	HBV DNA	Alt	Diagnosa
6	Agas	semarang tengah/semarang/20	0999	2	5	0	0	0	Pasien Terdiagnosa Hepatitis B

**Hasil Diagnosa Dengan Tabel 20 Data**

No	Nama Pasien	Alamat	TTL	Umur	No Telp	Hbe AG	HBV DNA	Alt	Diagnosa
6	Agas	semarang tengah/semarang/20	0999	2	5	0	0	0	Pasien Terdiagnosa Hepatitis B

Gambar 7. Tampilan Hasil Pengujian Data

Jika kita mempunyai data training set yang baru, maka kita bisa mempermudah perhitungan *fuzzy entropy* dan *gain*nya seperti pada Gambar 8. Dan untuk tampilan *database* terlihat pada Gambar 9.

**Entropi dan Gain Baru**

Input Parameter Entropi & Gain yang Baru

**Diagnosa**  
 sampel kelas 1 'Ya' =   
 sampel kelas 2 'Tidak' =

**Umur**  
 Umur Muda 'Ya' =  -> 'tidak' =   
 Umur Dewasa 'Ya' =  -> 'tidak' =   
 Umur Tua 'Ya' =  -> 'tidak' =

**HbeAG**  
 HbeAG Positif 'Ya' =  -> 'tidak' =   
 HbeAG Negatif 'Ya' =  -> 'tidak' =

**HBV-DNA**  
 HBV-DNA Rendah 'Ya' =  -> 'tidak' =   
 HBV-DNA Sedang 'Ya' =  -> 'tidak' =   
 HBV-DNA Tinggi 'Ya' =  -> 'tidak' =

**Alt**  
 Alt Rendah 'Ya' =  -> 'tidak' =   
 Alt Sedang 'Ya' =  -> 'tidak' =   
 Alt Tinggi 'Ya' =  -> 'tidak' =

Hasil

Gambar 8. Tampilan Untuk Menghitung Entropy Baru.

No Pasien	Nama Pasien	Alamat	TTL	Umur	No Telp	Hbe AG	HBV DNA	Alt
1	agus	lampar	semarang	31	123	1	3	0
2	rangga	iboyani	semarang, 1 januari 1992, 23	0825181	1	5	200	
3	ari	lampar tengah	beja	65	02999	2	5	150
4	rudi	semarang tengah kendal		45	0888	1	7	200
5	baei	semarang tengah kendal		30	08999	2	5	200
6	agus	semarang tengah semarang		20	0599	2	5	0

Gambar 9. Tampilan database.

### Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat diambil simpulan. Yang pertama, *Fuzzy ID3 (FID3)* dapat diterapkan

terhadap data hepatitis dengan ketentuan sebagai berikut; (a) Aturan – aturan *fuzzy* perhitungan *FID3* dari 10 data *training set* adalah (R1) IF HBV DNA Rendah Then Negatif Hepatitis B (R2) IF HBV DNA Sedang And Umur Muda Then Positif Hepatitis B (R3) IF HBV DNA Sedang And Umur Dewasa Then Negatif Hepatitis B (R4) IF HBV DNA Sedang And Umur Tua Then Positif Hepatitis B (R5) IF HBV DNA Tinggi Then Positif Hepatitis B. (b) Aturan – aturan *fuzzy* perhitungan *FID3* dari 15 data *training set* adalah (R1) IF HBV DNA Rendah Then Negatif Hepatitis B (R2) IF HBV DNA Sedang And Umur Muda Then negatif Hepatitis B (R3) IF HBV DNA Sedang And Umur Dewasa And HbeAg Positif Then Positif Hepatitis B (R4) IF HBV DNA Sedang And Umur Dewasa And HbeAg Negatif And Alt Rendah Then Positif Hepatitis B (R5) IF HBV DNA Sedang And Umur Dewasa And HbeAg Negatif And Alt sedang Then Positif Hepatitis B (R6) IF HBV DNA Sedang And Umur Dewasa And HbeAg Negatif And Alt Tinggi Then Negatif Hepatitis B (R7) IF HBV DNA Sedang And Umur Tua Then Positif Hepatitis B (R8) IF HBV DNA Tinggi Then Psitif Hepatitis B. (c) Aturan – aturan *fuzzy* perhitungan *FID3* dari 20 data *training set* adalah (R1) IF HBV DNA Rendah Then Negatif Hepatitis B (R2) IF HBV DNA Sedang And Alt Rendah Then negatif Hepatitis B (R3) IF HBV DNA Sedang And Alt Sedang Then Positif Hepatitis B (R4) IF HBV DNA Sedang And Alt Tinggi Then negatif Hepatitis B (R5) IF HBV DNA Tinggi Then Positif Hepatitis B.

Kedua, Hasil Penerapan *FID3* terhadap data hepatitis B adalah sebagai berikut: (a) Dari perhitungan ketiga *training set* tersebut, pembentukan *FID3* dengan 15 data *training set* yang paling akurat dengan 88,5%. (b) Pembentukan *Fuzzy ID3* sangat berpengaruh pada data *training set*nya. Semakin banyak dan semakin akurat datanya maka akan menaikan keakurasian *Fuzzy ID3*nya. (c) *Fuzzy ID3* dapat diimplementasikan dengan baik menggunakan *PHP MySQL* untuk mempermudah perhitungan.

### DAFTAR PUSTAKA

- Cahyono, S B. 2006. *Diagnosis dan Manajemen Hepatitis B Kronis*. <http://www.budilukmanto.org/index.php/seputar-hepatitis/80-seputar-hepatitis> diakses pada tanggal 14 November 2014.
- Elmande, Y & Pudjo, PW. 2012. Pemilihan *Criteria Splitting* dalam Algoritma *Iterative Dichotomiser 3 (ID3)* untuk Penentuan Kualitas Beras : Studi

- Kasus Pada Perum Bulog Divre Lampung. *Jurnal Telematika MKom Vol.4*
- Kusumadewi, S., Hartati, S., Harjoko, A. & Wardoyo, R. 2006. *Fuzzy Multi-Attribute Decision Making (FUZZY MADM)*. Yogyakarta : Graha Ilmu.
- Kusumadewi, S & Purnomo, H. 2003. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta : Graha Ilmu.
- Rahardjo, AS 2002. Aplikasi E-Commerce [www.komputeronline.com](http://www.komputeronline.com) Dengan Menggunakan MYSQL dan PHP4. *Jurnal Informatika Vol 3, No.2*. Universitas Kristen Petra.
- Ridwan, M., Suyono, H & Sarosa. 2013. Penerapan *Data Mining* Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier*. *Jurnal EECCIS Vol.7, No. 1*.
- No.1. Universitas Budi Luhur.
- Romansyah., S& Nurdiati. 2009. *Fuzzy Decision Tree dengan Algoritma ID3 pada Data Diabetes*. *Internetworking Indonesia Jurnal Vol. 1, No. 2*.
- Rosalina, I. 2012. Hubungan Polimorfisme Gen TLR 9 (RS5743836) dan TLR 2 (RS3804099 dan RS3804100) dengan Pembentukan Anti-HBS pada Anak Pascavaksinasi Hepatitis B. *Jurnal IJAS Vol.2, No.3*.
- Susilo, F. 2003. "*Himpunan dan Logika Kabur Serta Aplikasinya*". Yogyakarta : Graha Ilmu
- Welling, L & Laura, T. 2005. *PHP And MySQL Web Development*. Sams Publishing.