

IMPLEMENTASI ALGORITMA C4.5 UNTUK MENENTUKAN PENERIMA BEASISWA DI STT HARAPAN MEDAN

Rismayanti¹

¹Dosen Tetap Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan Medan
Jl. H.M Joni No.70 Medan

¹risma.stth@gmail.com

ABSTRACT

In educational institutions especially universities, there are many scholarships offered to the best and necessitous students. The scholarships is also given to the students in STT Harapan Medan, and one of the scholarship is Improving Academic Achievement Scholarship (PPA) which is given to outstanding students, in which grade point average (GPA) become a major criterion in determining the scholarship. C4.5 algorithm can help in determining the eligible students in getting scholarship in STTH Medan. C4.5 algorithm has main advantage that is it can produce a model like decision tree, which is easy to interpret for producing a decision rule. In analyzing and designing to determine the scholarship recipients, there are four attributes predictors used such as GPA, Semester, Parents income and number of Dependents, and there are two attribute predictions namely Accepted and Denied. The results of the implementation is done by using software Data Mining Rapid Miner, where the results will prove that the analysis by applying algorithm C4.5 is suitable with the implementation. The result of this writing is decision tree and rules to determine scholarship recipients for students of STT Harapan Medan.

Keyword: Algoritma C4.5, Decision Tree

INTISARI

Dalam lembaga pendidikan khususnya universitas dan sekolah tinggi banyak sekali beasiswa yang ditawarkan kepada mahasiswa yang berprestasi dan bagi mahasiswa yang kurang mampu. Pemberian bantuan belajar berupa beasiswa juga diberikan kepada mahasiswa di STT-Harapan Medan, salah satu beasiswa yang diberikan adalah Beasiswa Peningkatan Prestasi Akademik (PPA) yang diberikan kepada mahasiswa berprestasi, di mana IPK menjadi kriteria utama dalam pemberian beasiswa tersebut. Algoritma C4.5 dapat membantu dalam menentukan mahasiswa yang layak dan tidak layak menerima beasiswa dari STTH Medan. Algoritma C4.5 memiliki kelebihan utama yaitu dapat menghasilkan model berupa decision tree yang mudah diinterpretasikan untuk menghasilkan aturan keputusan. Dalam melakukan analisa dan perancangan untuk menentukan penerima beasiswa, maka ada 4 atribut prediktor yang digunakan yaitu IPK, Semester, Penghasilan Orang Tua dan Jumlah Tanggungan Orang Tua, sedangkan untuk atribut prediksi ada dua yaitu Diterima dan Ditolak. Hasil implementasi dilakukan dengan menggunakan software Data Mining Rapid Miner, di mana hasil nantinya akan membuktikan bahwa hasil analisa yang dilakukan dengan menerapkan algoritma C4.5 sesuai dengan hasil implementasi. Hasil yang diharapkan adalah sebuah decision tree dan rules untuk menentukan penerima beasiswa pada Mahasiswa di STT Harapan Medan.

Kata Kunci: Algoritma C4.5, Decision Tree.

I. PENDAHULUAN

Untuk menyelenggarakan pendidikan yang bermutu diperlukan biaya pendidikan yang cukup besar, oleh karena itu bagi setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan biaya pendidikan bagi mereka yang orang tuanya tidak mampu membiayai pendidikannya dan berhak mendapatkan beasiswa bagi mereka yang berprestasi. Pemberian bantuan belajar berupa beasiswa juga diberikan kepada mahasiswa di STT Harapan Medan.

Beasiswa yang diberikan antara lain beasiswa Peningkatan Prestasi Akademik (PPA) yang diberikan kepada mahasiswa berprestasi, beasiswa Bantuan Belajar Mahasiswa (BBM) yang diberikan kepada mahasiswa yang kurang mampu serta beasiswa yang

diberikan pada mahasiswa dimana IPK nya paling tinggi pada semester satu dan dua perjurusan masing-masing.

Algoritma C4.5 dapat membantu dalam menentukan mahasiswa yang diterima dan ditolak sebagai penerima beasiswa dari STT Harapan Medan. Algoritma C4.5 memiliki kelebihan utama yaitu dapat menghasilkan model berupa *tree* atau aturan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, dapat menangani atribut bertipe diskrit dan numerik. Pada algoritma C4.5, model yang dihasilkan oleh proses “belajar” dari data pelatihan berupa sebuah *decision tree*. *Decision tree* ini kemudian dapat dimanfaatkan untuk memprediksi kelas dari kasus yang baru.

Dalam menentukan penerima Beasiswa, ada 4 atribut prediktor yang digunakan yaitu IPK, Semes-

ter, Penghasilan Orang Tua dan Jumlah Tanggungan Orang Tua. Empat atribut prediktor tersebut merupakan kriteria yang digunakan STT Harapan Medan dalam menentukan penerima Beasiswa PPA.

II. TINJAUAN PUSTAKA

A) Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk *decision tree* berdasarkan *training data*. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal[4]. *Decision tree* adalah model prediksi menggunakan struktur *tree* atau struktur berhirarki. Konsep dari *decision tree* adalah mengubah data menjadi *decision tree* dan aturan-aturan keputusan (*decision rules*). Algoritma C4.5 merupakan pengembangan dari ID3.

Beberapa pengembangan yang dilakukan C4.5 adalah bisa mengatasi *missing value*, *continue data* dan *pruning*. Algoritma C4.5 mempunyai *input* berupa *training samples* dan *samples, training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya, sedangkan untuk *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data[6].

Algoritma C4.5 dan *decision tree* merupakan dua model yang tidak terpisahkan. Karena untuk membangun sebuah *decision tree*, dibutuhkan algoritma C4.5[2]. Di akhir tahun 1970 hingga di awal tahun 1980, J. Ross Quinlan seorang peneliti dibidang mesin pembelajaran membuat pengembangan sebuah *decision tree* yang dinamakan ID3 (*Iterative Dichotomizer*), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin dan PT. Stone, kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis *supervised learning* [2].

Serangkaian perbaikan yang dilakukan pada ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan berpengaruh untuk *decision tree* yaitu C4.5. Adapun perbaikannya adalah sebagai berikut:

- 1) Algoritma C4.5 menghitung *gain ratio* untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul atau *node*. Penggunaan *gain ratio* ini memperbaiki kelemahan dari ID3 yang menggunakan *information gain*.
- 2) Pemangkasan (*pruning*) dapat dilakukan pada saat pembangunan pohon(*tree*) ataupun pada saat proses pembangunan *tree* selesai.
- 3) Mampu menangani *continues attribute*.
- 4) Mampu menangani *missing data*.

- 5) Mampu membangkitkan *rule* dari sebuah *tree*.

Ada beberapa tahapan dalam membuat sebuah *decision tree* dengan algoritma C4.5 [2] yaitu:

- 1) Mempersiapkan *data training*. *Data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
- 2) Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan dipilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Di mana:

S = himpunan kasus

N = jumlah partisi S

P_i = proporsi S_i terhadap S

3. Menghitung nilai *Gain* menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (2)$$

Di mana:

S = himpunan kasus

A = fitur

N = jumlah partisi atribut A

$|S_i|$ = proporsi S_i terhadap S

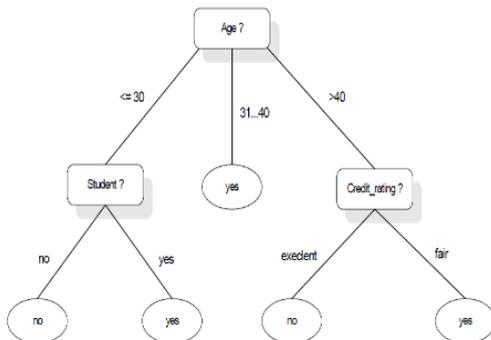
$|S|$ = jumlah kasus dalam S

4. Ulangi langkah ke-2 dan langkah ke-3 hingga semua *record* terpartisi
5. Proses partisi *decision tree* akan berhenti saat:
 - a) Semua *record* dalam simpul N mendapat kelas yang sama.
 - b) Tidak ada atribut didalam *record* yang dipartisi lagi.
 - c) Tidak ada *record* didalam cabang yang kosong.

B) Decision Tree

Salah satu metoda *Data Mining* yang umum digunakan adalah *decision tree*. *Decision tree* adalah struktur *flowchart* yang mempunyai *tree*(pohon), di mana setiap simpul *internal* menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas [7]. Alur pada *decision tree* ditelusuri dari simpul ke akar ke simpul daun yang memegang prediksi kelas. *Decision tree* adalah salah satu metode yang digunakan untuk pengklasifikasian dan prediksi karena memiliki kemudahan dalam interpretasi hasil [5].

Decision tree juga dapat digunakan untuk mengestimasi nilai dari variabel *continue* meskipun ada beberapa teknik yang lebih sesuai untuk kasus ini. Banyak algoritma yang dapat dipakai dalam pembentukan *decision tree*, antara lain *ID3*, *CART*, dan *C4.5*. *Decision tree* biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan, seperti terlihat pada gambar di bawah ini.



Gambar 1. *Decision Tree* Pembelian Komputer

Untuk mengklasifikasikan sampel yang tidak diketahui, nilai atribut dari sampel tersebut dites oleh *decision tree*. Suatu jalur ditelusuri dari *root* ke *node leaf* yang memiliki prediksi kelas untuk sampel tersebut. *Decision tree* dengan mudah dapat diubah dari aturan (*rule*) prediksi ke klasifikasi. Saat *decision tree* dibuat, kebanyakan dari cabang mungkin memperhatikan *noise* atau *outlier* pada *training data*, *tree pruning* berusaha untuk mengidentifikasi dan membuang cabang-cabang tersebut dengan tujuan memperbaiki akurasi klasifikasi pada data yang tidak kelihatan.

C) Beasiswa PPA

Beasiswa Peningkatan Prestasi Akademik (PPA) merupakan salah satu dari jenis beasiswa yang diberikan oleh pemerintah melalui pihak STTH kepada mahasiswa. Beasiswa PPA adalah beasiswa yang diberikan untuk peningkatan pemerataan dan kesempatan belajar bagi mahasiswa yang mengalami kesulitan membayar biaya pendidikannya sebagai akibat krisis ekonomi, terutama bagi mahasiswa yang berprestasi akademik. Adapun tujuan beasiswa Peningkatan Prestasi Akademik secara umum yaitu :

- 1) Meningkatkan pemerataan dan kesempatan belajar bagi mahasiswa yang mengalami kesulitan membayar pendidikan.
- 2) Mendorong dan mempertahankan semangat belajar mahasiswa agar dapat menyelesaikan studi pendidikan tepat waktunya.
- 3) Mendorong untuk meningkatkan prestasi akademik sehingga memacu peningkatan kualitas pendidikan.

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan berdasarkan kerangka kerja yang terdiri dari langkah-langkah berikut ini:

A) Identifikasi Masalah

Masalah pada penelitian ini bagaimana menentukan penerima beasiswa PPA di STT Harapan Medan. Dari masalah tersebut kemudian menentukan langkah yang tepat untuk memperoleh solusi yang terbaik.

B) Menganalisa Masalah

Merupakan langkah untuk dapat memahami masalah yang telah ditentukan dalam ruang lingkup atau batasannya.

C) Menentukan Tujuan

Tujuan yang akan dicapai dari penelitian ini membantu dalam menentukan calon penerima beasiswa pada Mahasiswa di lingkungan STTH Medan.

D) Mempelajari Literatur

Untuk mencapai tujuan penelitian, maka dipelajari literatur-literatur yang berkaitan dengan topik penelitian dan dapat dijadikan dasar atau sumber rujukan penelitian ini.

E) Mengumpulkan Data

Pada tahap pengumpulan data, penulis memperoleh data dengan metode pengisian angket oleh Mahasiswa STT Harapan Medan.

F) Merancang Model *Decision Tree* dengan Algoritma *C4.5*.

Pada tahap ini dilakukan proses yang meliputi 3 tahapan proses, tahap pertama yaitu *cleaning data*, *task-relevant data* yaitu melakukan seleksi data yang memiliki atribut yang relevan, dan tahap ketiga yaitu transformasi data. Data yang sudah ditransformasi, kemudian dianalisa menggunakan algoritma *C4.5* dengan menghitung *entropy* dan *gain* untuk menghasilkan *decision tree*.

G) Implementasi

Berdasarkan hasil rancangan yang diperoleh, kemudian di implementasikan dengan melakukan pengujian data menggunakan software Rapid Miner untuk menghasilkan *tree* yang sesuai dengan hasil analisis menggunakan algoritma *C4.5*.

IV. ANALISA DAN PERANCANGAN

H) Analisa Data

Untuk menentukan penerima beasiswa menggunakan metode algoritma C4.5 maka data yang akan digunakan dan dianalisa, yaitu data mahasiswa di Sekolah Tinggi Teknik Harapan (STTH). Di mana diperoleh dari pengisian angket oleh mahasiswa, yaitu: IPK, Semester, Penghasilan Orang Tua(PO) dan Jumlah Tanggungan Orang Tua(JTO). Ke empat atribut tersebut akan dijadikan atribut prediktor untuk menghasilkan atribut target, di mana atribut target tersebut menjadi *class output* untuk menentukan penerima beasiswa yaitu dibedakan menjadi 2 *class* (diterima dan ditolak).

I) Transformasi Data

Data yang akan ditransformasi adalah IPK, Jumlah Tanggungan Orang Tua (JTO), Penghasilan Orang Tua (PO) dan Daya Listrik. Data-data tersebut merupakan data yang akan dijadikan atribut *input* untuk merancang model *decision tree* menggunakan algoritma C4.5. Proses transformasi dilakukan dengan cara membuat klasifikasi pada masing-masing atribut prediktor, seperti terlihat pada Tabel 1.

Tabel 1. Keputusan Penentuan Penerima Beasiswa dari 20 sample Data Mahasiswa

Mahasiswa	Semester	IPK	PO	JTO	Hasil
M ₁	2	cum laude	tinggi	cukup	diterima
M ₂	2	Good	sedang	cukup	ditolak
M ₃	2	cum laude	tinggi	banyak	diterima
M ₄	2	cum laude	tinggi	banyak	diterima
M ₅	2	cum laude	tinggi	banyak	diterima
M ₁₅	4	very good	tinggi	banyak	diterima
M ₁₆	4	very good	tinggi	cukup	ditolak
M ₁₇	4	cum laude	tinggi	banyak	diterima
M ₁₈	4	Good	tinggi	banyak	ditolak
M ₁₉	4	cum laude	tinggi	banyak	diterima
M ₂₀	4	very good	tinggi	banyak	diterima
M ₂₁	4	very good	tinggi	cukup	ditolak
M ₂₂	4	cum laude	tinggi	banyak	diterima
M ₂₃	4	Good	tinggi	banyak	ditolak
M ₂₄	4	cum laude	tinggi	banyak	diterima
M ₆₇	6	Good	tinggi	banyak	ditolak
M ₆₈	6	cum laude	tinggi	cukup	diterima
M ₆₉	6	very good	sedang	banyak	diterima
M ₇₀	6	very good	sedang	cukup	ditolak
M ₈₃	6	very good	sedang	cukup	ditolak

Tabel 1 merupakan 20 sample data keputusan penentuan penerima beasiswa yang diambil dari secara acak dari data mahasiswa di STT Harapan Medan. Data tersebut nantinya akan digunakan untuk merancang model *decision tree*.

J) Merancang Decision Tree dengan Algoritma C4.5.

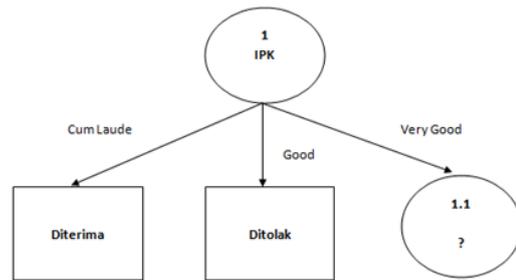
Berdasarkan Tabel 1, maka dilakukan proses perhitungan *entropy* (1) dan *gain* (2) untuk menentukan akar(*root*) dari pohon keputusan dalam membantu menentukan penerima beasiswa di STT Harapan Medan. Di bawah ini adalah hasil

perhitungan *entropy* dan *gain* pada node 1, pada Tabel 2.

Tabel 2. Hasil Perhitungan Entropy dan Gain pada Node 1.

Node	Atribut	Kelas	Jumlah Kasus (S)	Ditolak	Diterima	Entropy	Gain
1	Total		20	8	12	0,97095	
	Semester	2	5	1	4	0,72193	0,19445
		4	10	4	6	0,97095	
		6	5	3	2	0,44218	
	IPK						0,62612
		Cum Laude	9	0	9	0	
		Very Good	7	4	3	0,98523	
		Good	4	4	0	0	
	PO						0,09186
		Tinggi	16	5	11	0,89604	
		Sedang	4	3	1	0,81128	
	JTO						0,16228
		Tinggi	0	0	0	0	
		Cukup	7	5	2	0,86312	
	Banyak	13	3	10	0,77935		

Dari hasil tabel 2 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah IPK, yaitu 0,62612. Dengan demikian IPK dapat dijadikan *node* akar. Dari hasil perhitungan nilai *entropy* dan *gain* pada *node* 1, maka dapat digambarkan *decision tree* sementara seperti terlihat pada gambar 2.



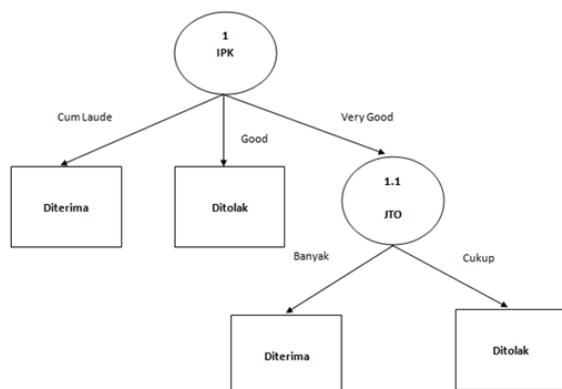
Gambar 2. Decision Tree Pada Node 1.

Untuk IPK yang bernilai *Very Good* belum mengklasifikasikan nilai menjadi satu keputusan, sehingga perlu dilakukan proses perhitungan *entropy* dan *gain* lebih lanjut pada node 1.1.

Tabel 3. Hasil Perhitungan Entropy dan Gain pada Node 1.1.

Node	Atribut	Kelas	Jumlah Kasus	Ditolak	Diterima	Entropy	Gain
1.1	IPK-	Very Good	7	4	3	0,98523	
		Semester					
		2	0	0	0	0	0,02024
		4	4	2	2	1	
		6	3	2	1	0,91830	
	PO						0,02024
		Tinggi	4	2	2	1	
		Sedang	3	2	1	0,91830	
	JTO						0,98523
		Tinggi	0	0	0	0	
Cukup		4	4	0	0		
	Banyak	3	0	3	0		

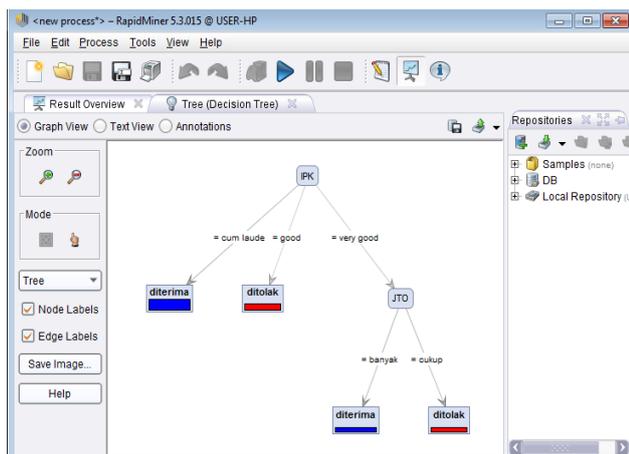
Berdasarkan Tabel 3 di atas maka dapat digambarkan *decision tree* untuk menentukan penerima beasiswa yang diterima dan ditolak pada Mahasiswa di STT Harapan Medan, terlihat pada Gambar 3.



Gambar 3. *Decision Tree* untuk Menentukan Penerima Beasiswa di STT Harapan Medan.

V. IMPLEMENTASI

Implementasi dilakukan menggunakan salah satu *software Data Mining* yaitu *Rapid Miner 5.3.015*. Semua atribut indikator *input* dan atribut tujuan disimpan dalam format *xlsx*, kemudian di *import* ke *software Rapid Miner 5.3.015* dan menghasilkan *decision tree* sebagai berikut:



Gambar 4. *Decision Tree* untuk Menentukan Penerima Beasiswa di STT Harapan Medan

Berdasarkan pengujian yang telah dilakukan, maka dapat disimpulkan bahwa hasil pengujian baik karena *decision tree* dan *rule* yang dihasilkan sesuai dengan pengujian secara manual dan pengujian menggunakan *Rapid Miner 5.3.015*.

Pada Gambar 4 dapat diketahui bahwa *node* akar dari *tree* adalah IPK dengan nilai Cum Laude, Very Good dan Good. Dari *node* akar dapat dijelaskan bahwa Mahasiswa yang memiliki IPK bernilai Cum Laude merupakan Mahasiswa yang diterima sebagai penerima beasiswa di STT Harapan Medan. Sedangkan untuk Mahasiswa yang memiliki IPK bernilai Good merupakan mahasiswa yang ditolak sebagai penerima beasiswa.

VI. PENUTUP

Setelah melakukan analisis, perancangan, implementasi beserta pengujian dalam menerapkan algoritma *C4.5* merancang *decision tree* untuk menentukan penerima beasiswa pada Mahasiswa STT Harapan Medan, maka didapat kesimpulan bahwa:

- 1) Dari hasil analisa data mahasiswa STT Harapan Medan, maka ada 4 atribut yang digunakan sebagai atribut prediktor yaitu Semester, IPK, Penghasilan Orang Tua(PO) dan Jumlah Tanggungan Orang Tua(JTO).
- 2) Hasil *Decision tree* penentuan penerima beasiswa menghasilkan keputusan bahwa mahasiswa yang 'diterima' sebagai penerima beasiswa adalah mahasiswa yang memiliki IPK bernilai Cum Laude dan IPK bernilai Very Good dengan Jumlah Tanggungan Orang Tua bernilai Banyak.
- 3) Hasil *Decision tree* penentuan penerima beasiswa menghasilkan keputusan bahwa mahasiswa yang 'ditolak' sebagai penerima beasiswa adalah mahasiswa yang memiliki IPK Good dan IPK bernilai Very Good dengan Jumlah Tanggungan Orang Tua bernilai Cukup.

VII. DAFTAR PUSTAKA

- [1] Abidin, Ae Zen Zainal. 2011. "Implementasi Algoritma C4.5 untuk Menentukan Tingkat Bahaya Tsunami". SemnasIF.
- [2] Andriani, Anik. 2013. "Sistem Prediksi Penyakit Diabetes Berbasis Decision Tree". Jurnal Informatika. Vol.1.No.1
- [3] Aribowo, Agus Sasmita 2013. "Metode Data Mining untuk Klasifikasi Kesetiaan Pelanggan Terhadap Merek Produk". SESINDO.
- [4] Ina, Wenefrida Tulit. 2013. "Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5". Jurnal Media Elektro. Vol.1.No.3.