

Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit

Larissa Navia Rani

Dosen Sistem Informasi UPI “YPTK” Padang
Jl. Raya Lubuk Begalung Padang - Sumatera Barat
larissa_navia_rani@upiypk.ac.id

Abstract - Customers are the most important assets in a bank. Possible payment many bad loans occur in various Bank. Classification of customers of Bank BRI is intended to facilitate the Bank in making decisions in the extension of credit. With the classification of customers in a bank, if a problem occurs with the same case the Bank live to see the rules (rule) that has been formed from the resulting decision tree. With this method of decision tree algorithm C4.5 desicion expected in the process of extracting information faster and optimized with larger data capacities, so that errors caused in decision making is minimized.

Keywords - Data Mining, Algoritma C4.5, DSS.

Intisari – Nasabah adalah aset terpenting dalam suatu Bank. Kemungkinan pembayaran kredit macet banyak terjadi diberbagai Bank. Klasifikasi nasabah Bank BRI ini bertujuan untuk memudahkan pihak Bank dalam membuat keputusan dalam hal perpanjangan kredit. Dengan adanya klasifikasi nasabah dalam suatu Bank, jika terjadi masalah dengan kasus yang sama pihak Bank tinggal melihat aturan-aturan (rule) yang telah terbentuk dari pohon keputusan yang dihasilkan. Dengan metode keputusan desicion tree menggunakan algoritma C4.5 diharapkan proses penggalian informasi lebih cepat dan optimal dengan kapasitas data yang lebih besar, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan.

Kata Kunci - Data Mining, Algoritma C4.5, DSS.

I. PENDAHULUAN

A. Latar Belakang Masalah

Perkembangan ilmu pengetahuan dan teknologi, mengakibatkan banyaknya terjadi perubahan di segala bidang, misalnya saja dalam dunia perbankan. Sebagai lembaga keuangan yang berfungsi menyimpan dana dari masyarakat dan menyalurkannya dalam bentuk pinjaman, Bank berusaha memberikan kemudahan dalam setiap layanan yang diberikannya, misalnya saja dalam layanan memberikan kredit bagi seorang debitur.

Namun dalam pemberian kredit tersebut, Bank bukan hanya sekedar memberikan pinjaman kepada seorang debitur, tetapi juga sekaligus menjadi sumber resiko operasi bisnis terbesar. Dengan demikian dalam pelaksanaannya bank harus memperhatikan asas-asas perkreditan yang sehat. Untuk mengurangi resiko tersebut, jaminan pemberian kredit

dalam arti keyakinan atas kemampuan dan kesanggupan debitur untuk melunasi hutangnya sesuai dengan yang diperjanjikan merupakan faktor yang sangat penting yang harus diperhatikan oleh bank.

Selain itu pada saat proses pengevaluasian permintaan kredit (misalnya kredit untuk sebuah perusahaan), seorang analisis kredit harus meneliti dan melakukan penilaian melalui suatu prosedur terhadap kondisi perusahaan calon debitur yang diperkirakan dapat mempengaruhi kemampuan mereka dalam memenuhi kewajiban kepada Bank. Salah satu tahapan dalam pengevaluasian dan penilaian tersebut adalah dengan melakukan pengklasifikasian nasabah berdasarkan kriteria-kriteria tertentu yang telah menjadi standar di setiap Bank. Salah satu metoda yang dapat digunakan untuk melakukan pengklasifikasian nasabah

tersebut adalah dengan melakukan analisis *Data Mining*.

B. Perumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan menjadi masalah-masalah sebagai berikut :

1. Bagaimana melakukan pengklasifikasian nasabah berdasarkan data karakteristik yang tersedia dengan algoritma C4.5 ?
2. Bagaimana membangun model perancangan *data mining* dalam menganalisa kriteria nasabah sebagai dasar pemberian kredit di Bank BRI Batam Unit Bengkong menggunakan algoritma C4.5 ?
3. Bagaimana menyajikan informasi klasifikasi nasabah sebagai dasar pemberian kredit menggunakan algoritma C4.5 ?

C. Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Mengetahui faktor-faktor yang diperlukan dalam melakukan pengklasifikasian nasabah sebagai dasar pemberian kredit.
2. Menentukan model yang dihasilkan dari teknik yang digunakan.
3. Mengimplementasikan algoritma C4.5 dalam melakukan pengklasifikasian nasabah pada proses pemberian kredit.
4. Menguji analisa penerapan *Data Mining* dengan teknik *classification* menggunakan algoritma C4.5 dalam melakukan pengklasifikasian nasabah pada proses pemberian kredit.

II. LANDASAN TEORI

A. Konsep Dasar Data Mining

Data Mining mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar [1]. Pramudiono dalam bukunya Kusri mengatakan *Data Mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [2].

1. Pengertian Klasifikasi

Klasifikasi adalah proses dari mencari suatu himpunan model (fungsi) yang dapat mendeskripsikan dan membedakan kelas-kelas data atau konsep-konsep, dengan tujuan dapat menggunakan model tersebut untuk memprediksi kelas dari suatu objek yang mana kelasnya belum diketahui.

2. Pengertian Algoritma C4.5

Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continuu data*, dan *pruning* [3]. Artikel jurnal di [4] menyebutkan algoritma C4.5 merupakan salah satu teknik *decision tree* yang sering digunakan, yang menghasilkan beberapa aturan-aturan dan sebuah pohon keputusan dengan tujuan untuk meningkatkan keakuratan dari prediksi yang sedang dilakukan, di samping itu algoritma C4.5 merupakan algoritma yang mudah dimengerti.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [2] :

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk tiap-tiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

3. Entropy dan Informasi Gain

Sebuah obyek yang diklasifikasikan dalam pohon harus dipes nilai *Entropy* -nya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut. *Entropy* (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. *Entropy* dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai *Entropy* maka akan semakin *Entropy* digunakan dalam

mengekstrak suatu kelas. *Entropy* digunakan untuk mengukur ketidakaslian S.sistem informasi atau disebut dengan *processing system*.

Besarnya *Entropy* pada ruang sampel S didefinisikan dengan :

$$Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus} \quad (1)$$

Di mana :

- S : ruang (data) sampel yang digunakan untuk pelatihan
- p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu .
- p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

Information gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan informasi *gain* tertinggi dipilih sebagai *test* atribut dari suatu *node* [5].

Gain (S,A) merupakan perolehan informasi dari atribut A *relative* terhadap *output* data S. Perolehan informasi didapat dari *output* data atau *variable dependent* S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A).

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

Di mana :

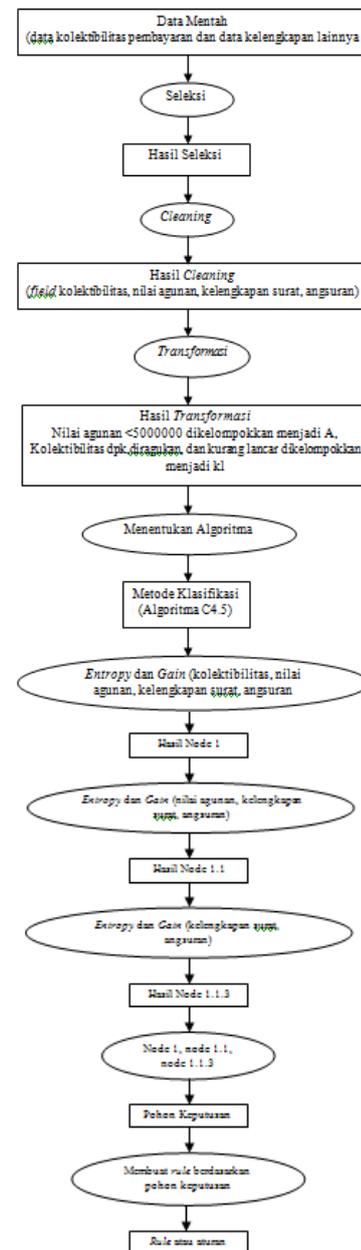
- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- $[S_i]$: jumlah kasus pada partisi ke-i
- $[S]$: jumlah kasus pada S

III. HASIL DAN PEMBAHASAN

A. Analisa Data Mining

Berdasarkan data yang dikumpulkan maka dilakukan proses seleksi, *cleaning* dan transformasi data sehingga ada 4 atribut yang diambil yaitu kolektibilitas, nilai agunan, kelengkapan surat, angsuran dengan kelas berpotensi dan tidak berpotensi. Setelah proses transformasi

selesai maka dilanjutkan ke proses *data mining* dengan teknik klasifikasi dan menggunakan algoritma C4.5, sehingga menghasilkan pohon keputusan dan aturan (*rule*). Berikut gambaran tentang langkah-langkah pengolahan *data mining* untuk menghasilkan *rule*.



Gambar 1 Tahapan *Data Mining*

B. Pengelompokan Data

Data-data yang dikumpulkan dikelompokkan berdasarkan atribut – atribut yang ada sehingga terlihat pada tabel I berikut.

TABEL I
PENGELOMPOKAN DATA BERDASARKNA
ATRIBUTNYA

Atribut	Jumlah	Berpotensi	Tidak Berpotensi
1	2	3	4
Total	42	30	12
<u>Kolektibilitas</u>			
LA	23	23	0
KL	17	7	10
M	2	0	2
<u>Nilai Agunan</u>			
A	2	2	0
B	5	4	1
C	4	4	0
D	9	6	3
E	5	2	3
F	2	1	1
G	4	2	2
H	7	6	1
I	4	4	0
<u>Kelengkapan Surat</u>			
L	28	26	2
TL	14	4	10
<u>Angsuran</u>			
J	3	3	0
K	7	6	1
L	1	1	0
M	8	5	3
N	4	1	3
O	5	3	2
P	2	1	1
Q	8	7	1
R	2	1	1
S	2	2	0

Data – data yang telah dikelompokkan dianalisa dengan persamaan matematik agar dapat menentukan *root* (akar). Untuk itu dilakukan proses perhitungan dan mencari nilai *Gain* dari masing – masing atribut yang ada. Jika sudah didapat nilai *Gain* dari masing – masing atribut maka diambil nilai *Gain* tertinggi sebagai *root* (akar).

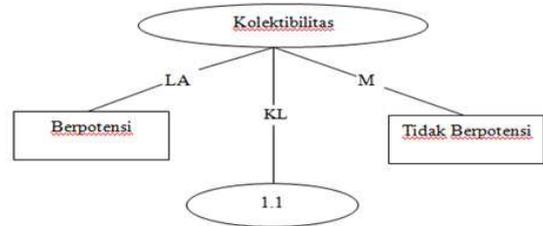
1. Perhitungan Node 1

TABEL II
HASIL PERHITUNGAN NODE 1

Atribut	Jumlah	Berpotensi	Tidak Berpotensi	Entropy	Gain
1	2	3	4	5	6
Total	42	30	12	0.863120569	
<u>Kolektibilitas</u>					0.467499071
LA	23	23	0	0	
KL	17	7	10	0.977417818	
M	2	0	2	0	
<u>Nilai Agunan</u>					0.22334044
A	2	2	0	0	
B	5	4	1	0.721928095	
C	4	4	0	0	
D	9	6	3	0.918295834	
E	5	2	3	0.970950594	
F	2	1	1	1	
G	4	2	2	1	
H	7	6	1	0.591672779	
I	4	4	0	0	
<u>Kelengkapan Surat</u>					0.327925495
L	28	26	2	0.371233327	
TL	14	4	10	0.863120569	
<u>Angsuran</u>					0.191083366
J	3	3	0	0	
K	7	6	1	0.591672779	
L	1	1	0	0	
M	8	5	3	0.954434003	
N	4	1	3	0.811278124	
O	5	3	2	0.970950594	
P	2	1	1	1	
Q	8	7	1	0.543564443	
R	2	1	1	1	
S	2	2	0	0	

Dari tabel II di atas dapat diketahui bahwa atribut yang memiliki *gain* tertinggi adalah atribut kolektibilitas yaitu 0.467499071. Dengan demikian, atribut kolektibilitas dapat menjadi node akar. Ada tiga nilai dari atribut kolektibilitas yaitu LA, KL, M. Atribut LA pada

kolektibilitas mengklasifikasikan kasus yaitu berpotensi sedangkan atribut M pada kolektibilitas mengklasifikasikan kasus yaitu tidak berpotensi. Dari hasil perhitungan tersebut dapat digambarkan pohon keputusan sementara tampak seperti gambar 2.



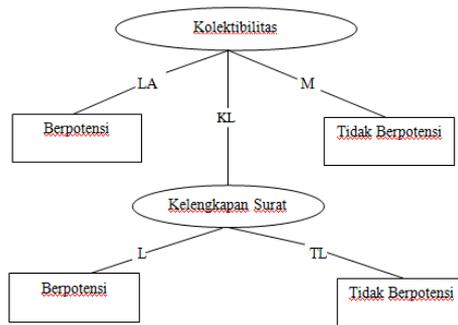
Gambar 2 Pohon Keputusan Hasil Perhitungan Node 1

2. Perhitungan Node 1.1

TABEL III
HASIL PERHITUNGAN NODE 1.1

Atribut	Jumlah	Berpotensi	Tidak Berpotensi	Entropy	Gain
Total	17	7	10	0.977417818	
<u>Nilai Agunan</u>					0.433587671
A	0	0	0	0	
B	2	1	1	0	
C	2	2	0	0	
D	4	2	2	1	
E	4	1	3	0.811278124	
F	1	0	1	0	
G	2	0	2	0	
H	2	1	1	1	
I	0	0	0	0	
<u>Kelengkapan Surat</u>					0.977417818
L	7	7	0	0	
TL	10	0	10	0	
<u>Angsuran</u>					0.265661025
J	0	0	0	0	
K	4	3	1	0.811278124	
L	0	0	0	0	
M	5	2	3	0.970950594	
N	2	1	1	1	
O	2	0	2	0	
P	1	0	1	0	
Q	2	1	1	1	
R	1	0	1	0	
S	0	0	0	0	

Dari tabel III di atas dapat diketahui bahwa atribut yang memiliki *gain* tertinggi adalah atribut kelengkapan surat yaitu 0.977417818. Dengan demikian atribut kelengkapan surat cabang keputusan terakhir dalam kasus ini. Ada dua nilai dari atribut kelengkapan surat yaitu L dan TL. Jadi variabel L pada kelengkapan surat mengklasifikasikan kasus yaitu berpotensi, sedangkan variabel TL pada kelengkapan surat mengklasifikasikan kasus yaitu tidak berpotensi. Dari hasil perhitungan tersebut dapat digambarkan pohon keputusan sementara yang tampak seperti gambar 3 berikut :



Gambar 3 Pohon Keputusan Hasil Perhitungan Node 1.1

Dari pohon keputusan akan terbentuk aturan – aturan dari penurunan data. Dari sampel calon nasabah yang akan melakukan perpanjangan kredit maka terbentuklah aturan atau *rule*.

Daftar aturan yang terbentuk adalah :

- a. Jika kolektibilitas lancar maka berpotensi.
- b. Jika kolektibilitas kurang lancar dan kelengkapan surat lengkap maka berpotensi.
- c. Jika kolektibilitas kurang lancar dan kelengkapan surat tidak lengkap maka tidak berpotensi.
- d. Jika kolektibilitas macet maka tidak berpotensi.

Berdasarkan hasil dari *rule* atau *knowledge* yang dihasilkan terdapat beberapa *rule* yang cukup sesuai dengan kejadian yang terjadi dalam klasifikasi calon nasabah yang akan memperpanjang kredit, nasabah yang memiliki riwayat pembayaran yang lancar dan sesuai dengan kriteria dasar pemberian kredit yang telah ditentukan maka nasabah tersebut berpotensi untuk diberikan kredit selanjutnya.

IV. IMPLEMENTASI

A. Data Yang Akan Diolah Dengan WEKA

Ketentuan-ketentuan untuk data yang akan diolah dengan WEKA adalah :

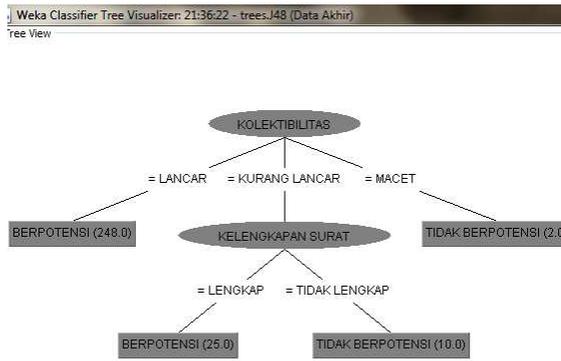
- a. Jika data yang disiapkan dalam bentuk Excel maka tidak boleh lebih dari satu *sheet*.
- b. Lalu simpan dengan format *CSV* (*comma separated value*).

Dalam penyimpanan tulis nama file dilanjutkan titik *CSV*. Contohnya : data akhir.CSV.

B. Analisa Decision Tree Menggunakan WEKA

Untuk mencari *knowledge* dengan algoritma C4.5 atau J.48 maka dilakukan langkah-langkah sebagai berikut :

1. Pada jendela utama WEKA klik tombol *explorer*. Lalu muncul jendela WEKA Explorer, kemudian pada *tab Preprocess* klik tombol *open file* untuk membuka data.
2. Selanjutnya keluar kotak dialog WEKA dan klik tombol *All* untuk mengeksekusi data.
3. Selanjutnya pada kotak dialog WEKA, aktifkan *tab classify*. Lalu klik tombol *Choose* pada *tab Classify* untuk mencari algoritma yang digunakan.
4. Pada menu *classify* pilih *Root Tree* dan klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan kriteria pembagian kelas dan kelas-kelas dengan menggunakan metode algoritma J48.
5. Setelah *classifiers*, pilihan-pilihan pengujian, dan *class* telah ditentukan, proses pembelajaran dapat dimulai dengan mengklik tombol *Start*, *user* dapat menghentikan proses ini sewaktu-waktu dengan tombol *Stop*. Saat *training* selesai, area *output classifiers* di sebelah kanan menampilkan *teks* yang menggambarkan hasil *training* dan pengujian. Sebuah *entry* baru saja muncul di kotak *Result list*.
6. Untuk membuat pohon keputusan maka klik kanan *trees-J.48* pada menu *result list* kemudian pilih *Visualize tree*.
7. Hasil *Visualize Tree* adalah hasil akhir dari pohon keputusan.



Gambar 4 Visualize Tree

Di mana *rule* yang dihasilkan pada gambar 4.1 adalah :

- a. IF kolektibilitas M THEN tidak berpotensi
- b. IF kolektibilitas KL AND kelengkapan surat L THEN berpotensi
- c. IF kolektibilitas KL AND kelengkapan surat TL THEN tidak berpotensi
- d. IF kolektibilitas LA THEN berpotensi.

C. Pembahasan Hasil Pengujian

Dari pohon keputusan dan *rule* yang terbentuk dimana status berpotensi dan tidak berpotensi dilihat dari atribut yang ada pada data awal yaitu data calon nasabah BRI BATAM dan hasilnya atribut kolektibilitas yang menjadi akar dan kelengkapan surat menjadi dahan atau node 1.1. Rincian dari jumlah calon nasabah yang berpotensi dan tidak berpotensi, terlihat pada tabel 4.1 berikut :

TABEL IV
RINCIAN JUMLAH MAHASISWA BERDASARKAN
RULE

No	Rule	Jumlah	
		Berpotensi	Tidak Berpotensi
1	Jika kolektibilitas LA maka berpotensi	248	
2	Jika kolektibilitas KL dan kelengkapan surat L maka berpotensi	25	
3	Jika kolektibilitas KL dan kelengkapan surat TL maka tidak berpotensi		10
4	Jika kolektibilitas macet maka tidak berpotensi		2

Berdasarkan tabel IV di atas dijelaskan bahwa setelah dilakukan pengujian dengan menggunakan *software* weka 3.7.4 untuk menganalisa kemungkinan pembayaran

kredit yang buruk yang dilihat dari proses berpotensi dan tidak berpotensi ketika mengajukan perpanjangan kredit ternyata hasil *rule* yang dihasilkan sama pada proses manual dalam pencarian *knowledge* dalam pohon keputusan.

V. KESIMPULAN

Dari uraian pada bab-bab yang sudah dibahas sebelumnya dapat ditarik kesimpulan :

1. Metode pohon keputusan (*decision tree*) yang diproses dengan *software* WEKA dapat mengidentifikasi kelayakan kredit dengan baik.
2. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang akan digunakan dalam menentukan sebuah klasifikasi juga sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan.
3. Sistem yang dibangun dapat membantu dalam mengklasifikasikan kriteria anggota nasabah yang pengajuan perpanjangan kreditnya berpotensi dan yang tidak berpotensi.
4. Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan (*decision tree*) maupun dalam bentuk aturan atau *rule If – Then* sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

Beberapa saran dari penulis untuk pengembangan tesis ini adalah :

1. Dalam menggunakan algoritma C4.5 untuk melakukan klasifikasi, harus dilakukan pemilihan variabel yang tepat agar hasil dari pohon keputusan lebih akurat atau terperinci.
2. Penelitian ini dapat dilanjutkan dengan menambahkan variabel-variabel yang memiliki hubungan dengan masalah pengajuan perpanjangan kredit agar nantinya tingkat pemberian

keputusan yang lebih baik dapat tercapai.

3. Penelitian ini dapat dilanjutkan dengan menggunakan metode-metode *data mining* lainnya untuk mendapatkan perbandingan metode yang lebih baik.

REFERENSI

- [1] Fatayat. dan Risanto. Joko. (2013). "Proses data mining Dalam Meningkatkan Sistem Pembelajaran Pada Pendidikan Sekolah Menengah Pertama." Kumpulan Makalah Seminar Semirata.
- [2] Kusrini dan Taufiq Luthfi, Emha. (2009). " Algoritma Data Mining." Yogyakarta : Andi.
- [3] Fadillah, Sarah. (2013). "Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer Dengan Menggunakan Algoritma C4.5." Pelita Informatika Budi Darma, Vol. 5, No. 3. 2301-9425.
- [4] Jantan, Hamidah., Razak Hamdan., Abdul dan Ali Othman, Zulaiha. (2010). "Human Talent Prediction in HRM using C4.5 Classification Algorithm." International Journal on Computer Science and Engineering, Vol. 02, No. 08. 2526-2534.
- [5] Ginting, Selvia Lorena Br., Zarman, Wendi, dan Hamidah, Ida. (2014) "Analisis dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik." Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNASTI). 1979-911X.