

# IMPLEMENTASI ALGORITMA DECISION TREE C4.5 UNTUK MENENTUKAN KELAS CALON SISWA DI LEMBAGA KURSUS BAHASA INGGRIS BERBASIS WEB

Uminingsih<sup>1</sup>, Ign.Suraya<sup>2</sup>, Isnanto Nugroho<sup>3</sup>

<sup>123</sup>Program Studi Sistem Komputer, Institut Sains & Teknologi AKPRIND Yogyakarta  
Email : umy\_bin@yahoo.com

## ABSTRACT

Difficulties in learning English in school led to the establishment of several English language course institutions. The number of classes / levels as well as the confidence factor of each student during the process. Less precisely in choosing this class that makes less effective use of time and money during the course. The implementation of C4.5 algorithm is very supportive for determining the class of prospective students. To facilitate the management of the selection system is made web-based using PHP, MySQL, and CSS. The attribute used is the number of answers to the question for classification and grammar abilities of prospective students with the results class are the Basic, Active, Pro, and Ultimate classes. This is due to advanced level institutions. From the classification process of the C4.5 algorithm, 12 rules are obtained with the grammar attribute as the main attribute. From 45 test data, there were 16 active classes, basic 15, pro 8, and ultimate class results 6. The accuracy obtained from this application was 91.11% with the amount of data predicted by 41 data.

**Keywords:** *Class, English, Classification, C4.5 Algorithm, Web*

## INTISARI

Kesulitan dalam belajar bahasa Inggris di sekolah memunculkan berdirinya beberapa lembaga kursus bahasa Inggris untuk membantunya. Banyaknya kelas/ jenjang serta factor percaya diri masing-masing siswa saat proses inialisasi membuat calon siswa mengalami kesulitan dalam memilih kelas yang sesuai. Kesalahan dalam memilih kelas ini yang membuat kurang efektifnya penggunaan waktu dan biaya selama kursus. Implementasi algoritma C4.5 sangat mendukung untuk menentukan kelas calon siswa. Untuk memudahkan pengelolaan sistem seleksi dibuat berbasis web dengan menggunakan PHP, MySQL, dan CSS. Atribut yang digunakan adalah jumlah jawaban pada pertanyaan untuk klasifikasi dan kemampuan *grammar* calon siswa dengan kelas hasil adalah kelas *Basic, Active, Pro, dan Ultimate*. Hal ini karena untuk lembaga kursus yang tingkat advance. Dari proses klasifikasi algoritma C4.5 diperoleh 12 *rules* dengan atribut *grammar* sebagai atribut utama. Dari 45 data uji diperoleh hasil kelas *active* sebanyak 16, basic 15, pro 8, dan ultimate sebanyak 6. Akurasi yang diperoleh dari aplikasi ini adalah sebesar 91,11% dengan jumlah data tepat prediksi sejumlah 41 data

**Kata Kunci :** Kelas, Bahasa Inggris, Klasifikasi, Algoritma C4.5, Web

## 1. PENDAHULUAN

Bahasa Inggris merupakan paspor untuk memasuki dunia luar. Banyak kendala orang mempelajari bahasa Inggris diantaranya, kurang efektifnya metode pengajaran dan kurangnya tenaga pengajar yang ahli dalam bidangnya. Fenomena inilah yang mendasari munculnya lembaga kursus bahasa Inggris. Pembelajaran di lembaga kursus bahasa Inggris untuk tingkat advance *conversation* ada 4 kelas/level yaitu kelas *basic, active, profesional, dan ultimate*. Pada penentuan level kelas biasanya dilakukan dengan memberikan beberapa tes dan setelah itu akan dilakukan wawancara untuk menguji kemampuan berbahasa Inggris. Permasalahannya terkadang siswa tidak percaya diri menjawab, sehingga banyak

sekali siswa yang salah masuk kelas. Hal ini akan mengurangi keefektifitasan waktu belajar siswa. Untuk mengatasi masalah tersebut perlulah dibuat sebuah sistem yang membantu calon siswa untuk mengetahui kelas yang cocok yang harus diambil oleh calon siswa dalam menentukan kelas dalam kursus bahasa Inggris salah satunya adalah dengan data mining.

Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Hermawati, F. A., 2013). Sedangkan, algoritma C4.5 yaitu metode pohon keputusan yang mengubah fakta yang

sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Algoritma C4.5 juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu (Suhartono, Suparlan. 2009). Pada algoritma ini dilakukan penghitungan entropy dan gain information untuk memperoleh node akar dan node lainnya. Proses minning berdasarkan Algoritma C4.5 dapat digunakan untuk melakukan seleksi kelas yang harus dipilih oleh calon siswa lembaga kursus bahasa Inggris yang mengacu pada data primer yang ada. Penulis memilih menggunakan algoritma C4.5 karena dapat melakukan prediksi dengan memberikan tingkat nilai akurasi yang tinggi .

Penelitian sebelumnya yang terkait penggunaan algoritma C.45, salah satunya adalah "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa" (Kamagi, D. H., dkk, 2014) yang membahas mengenai pendekatan teknik klasifikasi data mining dengan menggunakan algoritma C4.5 yang diterapkan dalam memprediksi tingkat kelulusan mahasiswa dengan empat kategori yaitu lulus cepat, lulus tepat, lulus terlambat dan drop out. Parameter yang digunakan adalah Indeks Prestasi Semester (IPS). Penelitian ini menghasilkan akurasi prediksi kelulusan mahasiswa dengan presentase 87.5% dari enam puluh data training dan empat puluh data testing. Hasil prediksi kelulusan dari aplikasi penelitian ini dapat membantu bagian program studi untuk mengetahui status kelulusan mahasiswa.

Pada penelitian ini digunakan algoritma C4.5 karena dapat melakukan prediksi dengan memberikan tingkat nilai akurasi yang tinggi untuk menentukan kelas calon siswa di Lembaga kursus bahasa Inggris. Tujuan dari penelitian ini berdasarkan uraian rumusan masalah di atas adalah sebagai berikut :a). Menentukan kelas calon siswa di lembaga kursus bahasa Inggris, b).Merancang bangun aplikasi penentu kelas calon siswa dengan algoritma C4.5 berbasis web. . Sedangkan level kelas dipilih: *basic*, *active*, *profesional*, dan *ultimate*.

## 2. METODE PENELITIAN

### 2.1. Basis Teori Perancangan

#### A. ALGORITMA C4.5

Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai

keluaran. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya. Algoritma C4.5 membangun pohon keputusan dari serangkaian data pelatihan yang mirip dengan Algoritma ID3, dengan menggunakan konsep entropi informasi. (Ginanjari, A. dkk., 2012)

Algoritma ini dapat menyelesaikan masalah secara sistematis dengan membentuk suatu *decision tree* dengan langkah-langkah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing masing record dari atribut.
3. Membagi kasus ke dalam cabang,
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang menghasilkan suatu keputusan yang sesuai.

Dalam memilih sebuah atribut menjadi akar, dilakukan perhitungan nilai dari atribut yang ada. Nilai gain yang paling tinggi dijadikan root pada pohon keputusan. Untuk menghitung nilai gain digunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1)$$

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S<sub>i</sub>| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Setelah mendapat nilai gain, ada satu hal lagi yang perlu dilakukan perhitungan yaitu mencari nilai entropy. Entropy digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan output dari atribut. Rumus dasar dari entropy tersebut adalah sebagai berikut :

$$Entropy(S) = \sum_{i=0}^n - P_i * \log_2 P_i \quad (2)$$

Keterangan :

S : himpunan kasus

n : jumlah partisi S

pi : proporsi dari S<sub>i</sub> terhadap S.

Setelah perhitungan selesai, maka dihasilkan pohon keputusan yang dapat diubah menjadi rules menggunakan kaidah IF-THEN-ELSE. Penjelasan rules tersebut adalah sebagai berikut:

#### 1. IF-THEN Rule

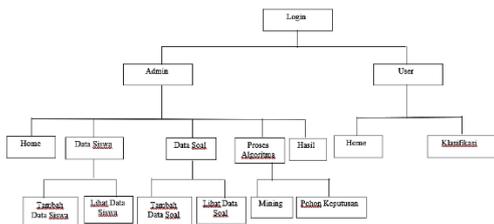
IF kondisi THEN Kesimpulan

Aturan ini terdiri dari dua bagian. Aturan yang (bagian IF) mengandung satu atau lebih

kondisi sekitar nilai prediktor atribut mana sebagai akibat aturan (THEN bagian) berisi prediksi tentang nilai atribut tujuan. Prediksi yang akurat dari nilai atribut tujuan akan meningkatkan proses pengambilan keputusan. IF-THEN aturan prediksi sangat populer di data mining. Mereka mewakili menemukan pengetahuan pada tingkat tinggi abstraksi. Aturan Induksi Metode memiliki potensi untuk menggunakan kasus diambil untuk prediksi.

### B.PERANCANGAN HIPO

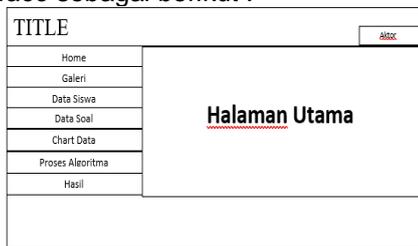
Perancangan HIPO untuk membangun sistem implementasi algoritma C4.5 untuk menentukan kelas calon siswa di lembaga kursus bahasa Inggris dilukiskan pada gambar : 1



Gambar 1. Perancangan HIPO sistem Aplikasi

### C.PERANCANGAN INTERFACE

Perancangan *interface* adalah bagian yang penting dalam aplikasi karena yang pertama kali dilihat ketika aplikasi dijalankan adalah *interface* aplikasi. Perancangan *interface* sebagai berikut :



Gambar 2. Perancangan Interface

### D. PERANCANGAN SISTEM

Pada bagian ini, akan dijelaskan tentang gambaran sistem yang akan dibangun menggunakan diagram pemodelan sistem UML. (Muslihudin, M., dkk., 2016)



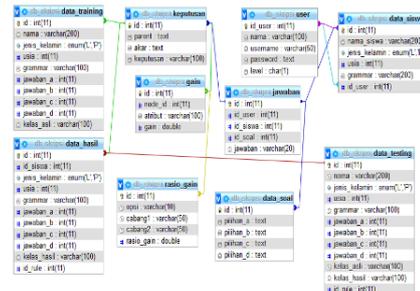
Gambar 3: Use Case Diagram

### Perancangan Database

Dalam pembuatan sistem klasifikasi ini basisdata bernama db\_skripsi yang terdapat 10 tabel yaitu tabel data\_hasil\_klasifikasi, data\_latih, data\_siswa, data\_uji, gain, jawaban\_kuisisioner, rasio\_gain, t\_keputusan, dan tabel users. Contoh strukturnya pada Tabel 2.

### RELASI ANTAR TABEL DATA

Pada Gambar .4. merupakan gambar relasi basisdata untuk aplikasi implementasi algoritma C4.5 yaitu database db\_skripsi dan mempunyai 10 tabel yang saling berelasi (Wahyudi, D. 2003).



Gambar 4. Relasi antar tabel

Tabel .1. Rancangan data\_hasil

No	Nama field	Tipe	Ukuran	Null	Keterangan
1	id	Int	11	No	Primary Key
2	id_siswa	int	11	No	id siswa
3	jenis_kelamin	enum	-	Yes	jenis kelamin user
4	usia	int	11	Yes	usia user
5	sekolah	varchar	100	Yes	asal sekolah user
6	jawaban_a	int	11	No	menampung jawaban a
7	jawaban_b	int	11	No	menampung jawaban b
8	jawaban_c	int	11	No	menampung jawaban c
9	jawaban_d	int	11	No	menampung jawaban d
10	kelas_hasil	varchar	100	No	hasil kelas user
11	id_rule	int	11	No	id rule yang cocok

## PERANCANGAN ALGORITMA *DECISION TREE* C4.5

Pada bagian ini akan disebutkan bagaimana proses jalannya algoritma *decision tree* C4.5. Berikut adalah langkah-langkah prosesnya (Nugroho, Y. S. 2014).  
*User* mengisi data pada aplikasi

1. Sistem menghitung total *entropy* sebelum dicari masing-masing *entropy class* dari masing-masing atribut
2. Sistem menghitung nilai *information gain* dan *gain ratio* dari masing-masing atribut.
3. Sistem menerima pemrosesan langkah ke-2 dan ke-3 hingga semua data terpartisi.
4. Jika :
  - a. Semua *record* didalam *node* N mendapat kelas yang sama.
  - b. Tidak ada atribut didalam *record* yang dapat dipartisi lagi.
  - c. Tidak ada *record* didalam *leaf* yang kosong.
 Maka : Proses partisi *decision tree* berhenti (*gain* = 0)  
 Selesai.

## HASIL DAN PEMBAHASAN Data Penelitian

Dalam pengambilan data dilakukan dua kali yaitu

- a). Data latih sebanyak 45 data adalah diambil dari peserta kursus pada 4 kelas yang sudah berjalan..Soal yang diberikan

untuk dijawabb adalah:1) bagaimana menurutnya atas kemampuan menguasai grammar, 2).Soal ABCD yg telah ditentukan, 3).mengisi pilihan kelas. Data ini setelah diolah (dicari entropi , gain dan dilakukan proses mining),maka akan menghasilkan pohon keputusan disertai rulanya. Dan ini nanti digunakan sebagai acuan untuk analisa data obyek(peserta kursus) yang akan diseleksi kelasnya(sebagai data uji) menghasilkan pohon keputusan disertai rulanya. Dan ini nanti digunakan sebagai acuan untuk analisa data objek (peserta kursus) yang akan diseleksi kelas nya (sebagai data uji).

- b). Untuk data uji juga diambil sebanyak 45 data diambil dari peserta pendaftar baru yang akan diseleksi kelasnya. Dengan menerapkan rule hasil dari data latih ke data uji maka akan diperoleh hasil seleksi kelas baru yang mendekati lebih tepat. Jadi hasil seleksi kelas baru kemungkinan bisa sesuai tapi juga bisa tidak sesuai

Data latih yang digunakan pada penelitian ini adalah berjumlah 45 data yang mana merupakan hasil pengisian data kuisioner oleh siswa yang sudah kursus dalam beberapa waktu dan menempati pada jenis kelas masing-masing. Data latih ditampilkan pada tabel 3.Selanjutnya data latih ini diklasifikasikan berdasarkan *field* datanya (Jumlah data, kelas Basic, kelas Active, kelas Pro, Kelas Ultimate). Untuk masing-masing jenis jawaban (Nugroho, Y. S. 2014).

Tabel 2. Data Siswa (sebagai data latih)

No	Nama	Jenis Kelamin	Usia	Grammar	Jawaban				Kelas Asli
					A	B	C	D	
1.	Mohammad Aziz .P	L	23	Baik	1	9	9	11	Ultimate
2.	Hajjar Nur Putriani	P	22	Baik	4	5	10	11	Ultimate
3.	Anggi Riyanto	L	23	Cukup	12	13	5	1	Active
4.	Angga Dewanti	P	26	Baik	1	4	3	22	Ultimate
5.	Bayu Bagus	L	23	Baik	13	14	2	1	Active
6.	Anggi Pratiwi	P	22	Cukup	10	10	6	4	Basic
7.	Muhammad Iqbal	L	23	Cukup	14	10	5	1	Basic
8.	Galih Firman	L	22	Cukup	18	8	2	2	Basic
9.	Miming Suryani	P	23	Baik	5	8	15	2	Pro
10.	Dhani Ramadhan	L	23	Baik	6	3	17	4	Pro
11.	Rizkhi Aswariyan	L	22	Baik	1	12	16	1	Pro
12.	Maharani R	P	23	Cukup	15	11	2	2	Basic
13.	Valentino Ibanez	L	23	Baik	10	14	4	2	Active
14.	Fitrianingrum	P	24	Baik	5	5	14	6	Pro
15.	Avra	P	24	Cukup	10	14	4	2	Active
16.	M Riyadi	L	24	Cukup	14	9	5	2	Basic
17.	Merrysa	P	23	Cukup	13	12	2	3	Active

18.	Siska Candra	P	23	Baik	7	6	14	3	Pro
19.	Mahardika Cahaya	L	22	Baik	2	2	11	15	Ultimate
20.	Hasan	L	21	Cukup	11	17	2	0	Active
21.	Eva Widi	P	23	Cukup	17	9	3	1	Basic
22.	Ira Musyarafa	P	24	Baik	3	2	22	4	Pro
23.	Rodhiatam Miftah	P	23	Baik	2	6	10	12	Ultimate
24.	Bagus Ardiyanto	L	25	Cukup	22	5	2	1	Basic
25.	Kiki	P	20	Baik	4	6	17	3	Pro
26.	Hani Irawati	P	20	Baik	1	4	6	19	Ultimate
27.	Elisabeth Amanda	P	23	Baik	1	3	12	14	Ultimate
28.	Panji	L	20	Cukup	6	19	4	1	Active
29.	Sri Mulyani	P	24	Baik	0	1	8	21	Ultimate
30.	Reni Wijayanti	P	26	Baik	3	6	15	6	Pro
31.	Susiana	P	23	Cukup	11	17	1	1	Active
32.	Yusia	P	22	Cukup	21	7	1	1	Basic
33.	Santi Ernawati	P	23	Cukup	18	6	4	2	Basic
34.	Salvatore Dimas	L	24	Baik	1	1	10	18	Ultimate
35.	Endicha	L	24	Cukup	7	16	4	3	Active
36.	Sugito	L	21	Cukup	4	15	7	4	Active
37.	Syahrul Gunawan	L	19	Cukup	15	12	2	1	Basic
38.	Muhammad Affif	L	24	Baik	2	2	9	17	Ultimate
39.	Thomas Nugroho	L	23	Baik	1	4	22	3	Pro
40.	Tian	L	19	Cukup	22	5	2	1	Basic
41.	Erika	P	23	Cukup	12	14	3	1	Active
42.	Arditya	L	22	Cukup	14	12	3	1	Basic
43.	Ayi	P	21	Baik	2	5	10	13	Ultimate
44.	Nur Seto Aji	L	24	Baik	4	6	9	11	Ultimate
45.	Eka Wulandari	L	22	Baik	2	7	13	8	Pro

Tabel .3 Klasifikasi Grammar

Atribut	Jumlah Data	Kelas Basic	Kelas Active	Kelas Pro	Kelas Ultimate
Grammar = baik	25	1	2	10	12
Grammar = cukup	20	11	9	0	0

Tabel 4. Klasifikasi Jawaban A

Atribut	Jumlah Data	Kelas Basic	Kelas Active	Kelas Pro	Kelas Ultimate
Jawaban A <= 10	28	1	5	10	12
Jawaban A > 10	17	11	6	0	0

Dengan cara yang sama dilakukan sampai pada Jawaban D

Setelah data diklasifikasikan langkah pertama dalam proses algoritma C 4.5 adalah menghitung total *entropy* yaitu dengan menerapkan Rumus (2) maka sebagai contoh :

$$Entropy(Total) = \left(-\frac{12}{45} * \log_2\left(\frac{12}{45}\right)\right) + \left(-\frac{11}{45} * \log_2\left(\frac{11}{45}\right)\right) + \left(-\frac{10}{45} * \log_2\left(\frac{10}{45}\right)\right) + \left(-\frac{12}{45} * \log_2\left(\frac{12}{45}\right)\right) = 1.998947$$

$$Entropy(Total) = 1.998947$$

Dengan cara yang sama dilakukan sampai ke jawaban D dan grammar serta disesuaikan dengan input data yang masuk dihitung semuanya.

Langkah selanjutnya adalah perhitungan untuk mencari nilai *Gain* menggunakan rumus persamaan(.1 )yang nantinya akan digunakan sebagai *root node* :, sebagai contoh

$$Gain (Total, A = 15) = 1.998947 - \left(\frac{39}{45} * 1.957\right) + \left(\frac{6}{45} * 0\right) = 0.3$$

$$Gain(Total, grammar) = 1.998947 - \left(\frac{25}{45} * 1.514\right) + \left(\frac{20}{45} * 0.993\right) = 0.714$$

Hasil perhitungan entropy dan gain ditunjukkan pada tabel 5:

Tabel. 5 Hasil Perhitungan gain dan entropy

Node	Atribut	Jumlah Kasus (S)	Basic (S1)	Active (S2)	Pro (S3)	Ultimate (S4)	Entropy	Gain	
1	Total Grammar	45	12	11	10	12	1.996		
	Jawaban A	A ≤ 10	28	1	5	10	12	1.67	0.603
		A > 10	17	11	6	0	0	0.993	
		A ≤ 15	39	6	11	10	12	1.957	0.3
		A > 15	6	6	0	0	0	0	
		A ≤ 20	42	9	11	10	12	1.992	0.137
		A > 20	3	3	0	0	0	0	
		A ≤ 25	45	12	11	10	12	1.996	0
		A > 25	0	0	0	0	0	0	
	Jawaban B	B ≤ 10	30	9	0	9	12	1.571	0.598
		B > 10	15	3	11	1	0	1.053	
		B ≤ 15	41	12	7	10	12	1.97	0.201
		B > 15	4	0	4	0	0	0	
		B ≤ 20	45	12	11	10	12	1.996	0
		B > 20	0	0	0	0	0	0	
		B < 25	45	12	11	10	12	1.996	0
		B > 25	0	0	0	0	0	0	
	Jawaban C	C ≤ 10	33	12	11	0	10	1.581	0.663
		C > 10	12	0	0	10	2	0.65	
		C ≤ 15	40	12	11	6	11	1.956	0.177
		C > 15	5	0	0	4	1	0.772	
		C ≤ 20	43	12	11	8	12	1.982	0.305
		C > 20	12	0	0	10	2	0.65	
		C ≤ 25	45	12	11	10	12	1.996	0
		C > 25	0	0	0	0	0	0	
	Jawaban D	D ≤ 10	33	12	11	9	1	1.723	0.622
		D > 10	12	0	0	1	11	0.414	
		D ≤ 15	40	12	11	10	7	1.973	0.242
		D > 15	5	0	0	0	5	0	
		D ≤ 20	43	12	11	10	10	1.996	0.089
		D > 20	2	0	0	0	2	0	
		D ≤ 25	45	12	11	10	12	1.996	0
		D > 25	0	0	0	0	0	0	
	Grammar	Baik	25	1	2	10	12	1.514	0.714
		Cukup	20	11	9	0	0	0.993	

Dari hasil perhitungan Gain diperoleh Gain tertinggi adalah grammar yaitu 0.714 sehingga grammar menjadi *root node*. *Root node* digunakan sebagai *rule* utama dalam pohon keputusan. Kemudian untuk memperoleh keputusan *leaf* harus dilakukan perhitungan kembali hingga *node* akhir menghasilkan gain sebesar 0.

#### Proses Pembentukan Pohon Keputusan Algoritma C4.5

Dengan mengambil data latihan sebanyak 45 buah yang telah diklasifikasikan kemudian dihitung *entropy* keseluruhannya,

dari data latihan yang ada diperoleh hasil seperti pada tabel 6

Tabel .6 Perhitungan Uji Awal Data

Jumlah Data	45
Jumlah <i>Basic</i>	12
Jumlah <i>Active</i>	11
Jumlah <i>Pro</i>	10
Jumlah <i>Ultimate</i>	12
<i>Entropy</i> Keseluruhan	1.996

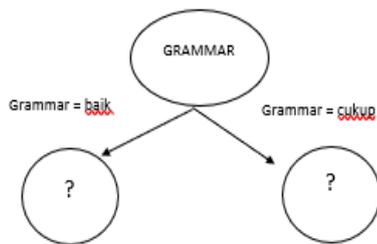
Kemudian data latihan yang sudah di klasifikasikan dihitung *gain* dan *entropy*nya menggunakan rumus(.1) dan rumus(2)

sehingga diperoleh hasil (contoh sebagian cuplikan hasil perhitungan gain dan entropy dilukiskan pada Gambar 5 :

Nilai Atribut	Jumlah data	Jumlah Basic	Jumlah Active	Jumlah Pro	Jumlah Ultimate	Entropy	Gain
grammar=baik	25	1	2	10	12	1.514	
grammar=cukup	29	11	9	0	0	0.883	0.714
jabatan_p=10	25	1	5	10	12	1.517	
jabatan_p=15	17	11	6	0	0	0.877	0.633
jabatan_b=10	29	9	11	10	12	1.587	
jabatan_b=15	9	9	9	0	0	1	0.3
jabatan_u=10	42	9	11	10	12	1.882	
jabatan_u=15	3	3	9	0	0	1	0.187
jabatan_c=10	45	12	11	10	12	1.906	
jabatan_c=15	9	9	9	0	0	1	0
jabatan_s=10	29	9	9	9	12	1.571	
jabatan_s=15	15	5	11	1	0	0.883	0.595

Gambar 5. Sebagian hasil cuplikan perhitungan Gain dan Entropy

Dari hasil gambar diatas(sampai 45 data) dapat diketahui bahwa atribut dengan Gain tertinggi adalah Grammar, yaitu 0.714 sehingga atribut Grammar menjadi *root node* dan harus dilakukan perhitungan kembali untuk mencari keputusan hingga perhitungan selesai atau berhenti (bila hasil iterasi mencapai 0) .



Gambar 6. .Gambar keputusan perhitungan pertama

Kemudian menghitung kembali jumlah kasus untuk Grammar = baik, jumlah kasus untuk keputusan kelas *basic*, *active*, *pro*, *ultimate* dan *entropy* dari data latih hingga proses *mining* selesai. Hasil iterasi adalah menentukan cabang .Proses iterasi ditampilkan pada gambar 7

```

=====
cabang 1
(grammar=baik)
Jumlah data = 25
Jumlah Basic = 1
Jumlah Active = 2
Jumlah Pro = 10
Jumlah Ultimate = 12
Entropy Ad = 1.514
  
```

Nilai Atribut	Jumlah data	Jumlah Basic	Jumlah Active	Jumlah Pro	Jumlah Ultimate	Entropy	Gain
grammar=baik	25	1	2	10	12	1.514	
grammar=cukup	0	0	0	0	0	0	0
jabatan_p=10	25	0	1	10	12	1.209	
jabatan_p=15	2	1	1	0	0	1	0.322
jabatan_b=10	29	0	2	10	12	1.325	
jabatan_b=15	1	1	0	0	0	0	0.242

Gambar 7. Mencari Gain dan Entropy Perhitungan ke 2

Kemudian dilakukan proses *mining* kembali dengan *root node* C=10. Atribut C=10 dipilih karena pada iterasi sebelumnya memiliki gain tertinggi. Selanjutnya dilakukan perhitungan gain dan entropi perhitungan ke 3 dilukiskan pada Gambar 8.

```

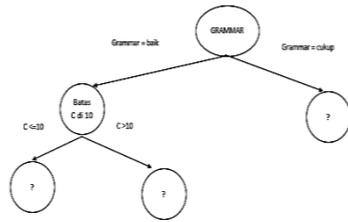
=====
cabang 1
(grammar=baik) AND (jabatan_p=10)
Jumlah data = 13
Jumlah Basic = 1
Jumlah Active = 2
Jumlah Pro = 0
Jumlah Ultimate = 10
Entropy Ad = 0.991
  
```

Nilai Atribut	Jumlah data	Jumlah Basic	Jumlah Active	Jumlah Pro	Jumlah Ultimate	Entropy	Gain
grammar=baik	13	1	2	0	10	0.991	
grammar=cukup	0	0	0	0	0	0	0
jabatan_p=10	13	0	1	0	10	0.459	
jabatan_p=15	2	1	1	0	0	1	0.466
jabatan_b=10	12	0	2	0	10	0.65	
jabatan_b=15	1	1	0	0	0	0	0.291
jabatan_u=10	12	0	2	0	10	0.65	
jabatan_u=15	1	1	0	0	0	0	0.291

Gambar 8. Mencari Gain dan Entropy Perhitungan 3

Kemudian dilakukan proses *mining* kembali dengan *root node* D=10. Atribut D=10 dipilih karena pada iterasi sebelumnya memiliki gain tertinggi. Pohon keputusan yang terbentuk dari tahap ini ditunjukkan pada gambar 9.

PERHITUNGAN 2



Gambar : 9 Pohon keputusan yang dihasilkan tahap perhitungan 2

Hasil analisa pohon keputusan sampai pada tahap tersebut dilukiskan pada gambar10..

```

=====
cabang 1
(grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d<=10) AND (jawaban_a<=15)

LEAF |Keputusan = Active
=====
cabang 2
(grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d<=10) AND (jawaban_a>15)

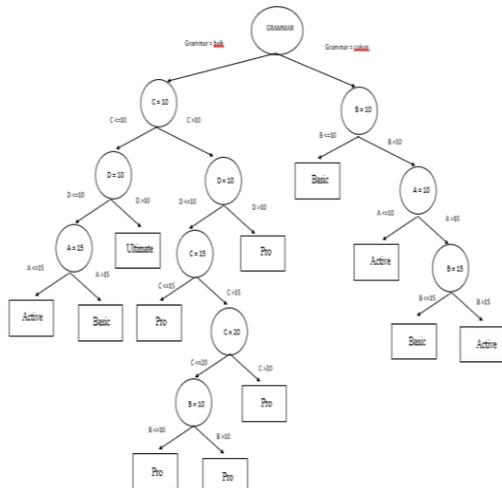
LEAF |Keputusan = Basic
=====
cabang 2
(grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d>10)

LEAF |Keputusan = Ultimate
=====

```

Gambar 10. Hasil analisis rule proses mining data latih.

Dengan cara yang sama proses mining dilakukan sampai menghasilkan 0 untuk seluruh data .[ Hermawati, F. A] .Seperti dilukiskan pada gambar 11.



Gambar 11 .Hasil pohon keputusan algoritma C-45 dari data latih

Dari pohon keputusan proses mining yang dilakukan menghasilkan 12 Rule. Hal ini terlihat pada gambar 12

Jumlah rule : 12

Id	Aturan
1	IF (grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d<=10) AND (jawaban_a<=15) THEN Label = Active
2	IF (grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d<=10) AND (jawaban_a>15) THEN Label = Basic
3	IF (grammar="Baik") AND (jawaban_c<=10) AND (jawaban_d>10) THEN Label = Ultimate
4	IF (grammar="Baik") AND (jawaban_c>10) AND (jawaban_d<=10) AND (jawaban_a<=15) THEN Label = Pro
5	IF (grammar="Baik") AND (jawaban_c>10) AND (jawaban_d<=10) AND (jawaban_a>15) AND (jawaban_b<=20) AND (jawaban_b>=10) THEN Label = Pro
6	IF (grammar="Baik") AND (jawaban_c>10) AND (jawaban_d<=10) AND (jawaban_a>15) AND (jawaban_b<=20) AND (jawaban_b>=10) THEN Label = Pro
7	IF (grammar="Baik") AND (jawaban_c>10) AND (jawaban_d<=10) AND (jawaban_a>15) AND (jawaban_b>20) THEN Label = Pro
8	IF (grammar="Baik") AND (jawaban_c>10) AND (jawaban_d>10) THEN Label = Pro
9	IF (grammar="Cukup") AND (jawaban_b<=10) THEN Label = Basic
10	IF (grammar="Cukup") AND (jawaban_b>10) AND (jawaban_c<=10) THEN Label = Active
11	IF (grammar="Cukup") AND (jawaban_b>10) AND (jawaban_c>10) AND (jawaban_d>15) THEN Label = Basic
12	IF (grammar="Cukup") AND (jawaban_b>10) AND (jawaban_c>10) AND (jawaban_d>15) THEN Label = Active

Gambar 12. Rule sebagai hasil Interpretasi Pohon Keputusan

PENGUJIAN ALGORITMA

Dari hasil penelitian ini untuk mengetahui ketepatan dalam menentukan metode maupun perancangan perlu dilakukan pengujian. b) pengujian keberhasilan hasil perancangan sistem menggunakan web.Hal tersebut dapat diterangkan berikut:

a. PENGUJIAN ALGORITMA C.45

Langkah-langkah pengujian:

1. Mengambil data uji sebanyak 45
2. Data dikenakan pada program mining,dan dianalis dengan rule
3. Membandingkan dengan hasil kelas asli( yang ditulis sendiri oleh siswa)

Hasil klasifikasi dengan penerapan rule sebagai contoh cuplikan beberapa data dilukiskan pada gambar 13

Hasil Klasifikasi

Cuplik

Jumlah data: 45

No	Nama	Jenis Kelamin	Usia	Grammar	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Kelas Hasil	Id rule
1	ekstrem10	P	22	Baik	15	15	0	0	Active	3
2	Siswa9	L	22	Cukup	7	22	1	0	Active	10
3	Siswa7	P	23	Cukup	9	17	3	1	Active	10
4	Siswa5	L	22	Baik	5	16	7	2	Active	1
5	Siswa13	P	23	Cukup	5	22	2	1	Active	10
6	Siswa17	L	23	Baik	6	18	5	1	Active	1
7	Siswa2	L	22	Baik	7	18	4	1	Active	1
8	Siswa1	L	22	Baik	7	12	7	4	Active	1

Gambar 13..Cuplikan hasil penerapan rule ke data uji

.Selanjutnya dilakukan uji akurasi: yaitu dengan membandingkan dengan hasil data asli(pilihan siswa sendiri). Dari hasil perbandingan dapat digunakan untuk uji ketepatan hasil dari algoritma *Decision Tree* C4.5 maka dihitung akurasi. Untuk menghitung akurasi digunakan rumus:

$$Akurasi = \frac{Total\ Prediksi\ Benar}{Total\ Prediksi} \times 100 \dots \dots (2.3)$$

Contoh Cuplikan sebagian hasil uji akurasi pada gambar 14

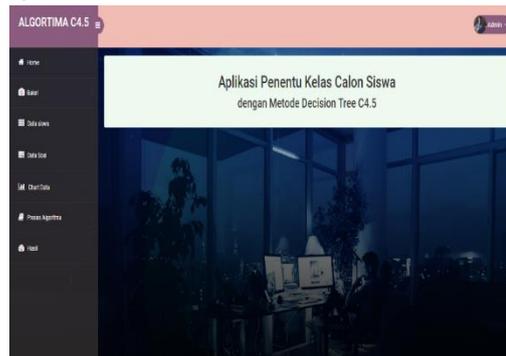
41	Soal18	L	23	Baik	15	9	5	1	Active	Active	1	benar
42	Soal19	P	20	Cukup	2	11	11	6	Active	Active	10	benar
43	Soal20	P	22	Cukup	17	9	3	1	Basic	Basic	9	benar
44	Soal21	L	22	Baik	12	11	4	3	Active	Active	1	benar
45	Soal22	P	23	Cukup	19	7	2	2	Basic	Basic	9	benar

Jumlah prediksi: 45  
 Jumlah tepat: 41  
 Jumlah tidak tepat: 4  
**AKURASI = 91.11 %**  
**LAJU ERROR = 8.89 %**

Gambar 14. Cuplikan hasil uji akurasi data uji.

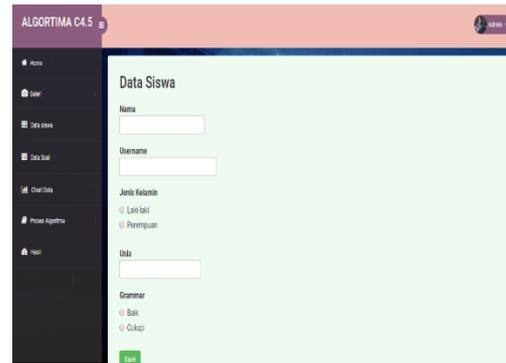
#### b).PENGUJIAN PERANCANGAN SISTEM

Pengujian sistem yang dilakukan ialah dengan menggunakan metode *blackbox testing*, adalah salah satu metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas, khususnya pada input dan output aplikasi. Hasil pengujian sesuai dengan perancangan an sistem adalah halaman utama admin. Halaman ini memuat informasi data soal serta hasil proses algoritma C4.5 terdiri dari proses mining dan penentuan pohon keputusan., seperti dilukiskan pada gambar 15.



Gambar 15. Halaman utama admin.

Selanjutnya untuk mengisikan data siswa oleh admin dilukiskan pada gambar 16.



Gambar 16. Halaman data siswa.

#### Kelebihan Sistem

Kelebihan yang dimiliki oleh sistem yang diteliti, yaitu

1. Bentuk implementasi dari sistem mudah dipahami dan mudah diakses.
2. Tampilan mudah dimengerti oleh pengguna awam
3. Tampilan mudah dimengerti oleh pengguna awam
4. User dimudahkan dalam mengklasifikasi kelas.
5. Pembagian aktivitas antar user jelas.
6. Mudah dalam perawatan dan pengembangan.
7. *Rule* yang dihasilkan dinamis berdasarkan data uji

#### KESIMPULAN

1. Implementasi algoritma C4.5 ini menghasilkan 12 *rules* dari hasil pengolahan data uji sebanyak 45 data dengan iterasi perulangan sebanyak 12 kali.
2. Atribut grammar merupakan atribut yang paling berpengaruh pada aplikasi Implementasi algoritma C4.5 ini karena grammar merupakan atribut yang menghasilkan *gain* tertinggi sehingga grammar menjadi *root node* pada pohon keputusan.
3. Dengan sistem ini untuk menentukan kelas calon siswa di lembaga kursus bahasa Inggris bisa diakses melalui web dengan cara membuka localhost/web pad
4. Hasil klasifikasi siswa sebanyak 45 data didominasi oleh kelas *active* sejumlah 8, dan *ultimate* sejumlah 6 data. kemudian dihasilkan *basic* sejumlah 15 data, *prov* sejumlah 5 data
5. Aplikasi klasifikasi ini menghasilkan akurasi sebesar 91.11% dan laju eror sebesar 8.89% dengan data prediksi tepat

adalah 41 dan prediksi tidak tepat adalah 4 dari 45 data uji.

#### **UCAPAN TERIMAKASIH**

Ucapan Terimakasih kepada IST Akprind yang telah membantu dalam pendanaan penelitian ini.

#### **DAFTAR PUSTAKA**

- Apriliyani, R. (2017). Pengaruh dan harga bukti fisik terhadap kepuasan dan dampaknya pada loyalitas peserta didik (studi kasus terhadap peserta didik yang kursus di EPLC Bandung).
- Ginanjari, A., Mabrur, & Lubis, R. (2012). Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit. *Komputa Edisi 1, Vol 1*.
- Hermawati, F. A. (2013). Data Mining. Penerbit Andi. Yogyakarta:
- Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi

Tingkat Kelulusan Mahasiswa. *Ultimatics Vol VI, No.1*.

- Muslihudin, M., & Oktafianto. (2016). *Analisis dan Perancangan Sistem Informasi Menggunakan Model Terstruktur dan UML*. Yogyakarta: Penerbit Andi.
- Muslihudin, M., & Oktafianto. (2016). *Analisis dan Perancangan Sistem Informasi Menggunakan Model Terstruktur dan UML*. Yogyakarta: Penerbit Andi.
- Nugroho, B. (2004). *Aplikasi Pemrograman Web Dinamis dengan PHP dan MySQL*. Yogyakarta: Gava Media.
- Nugroho, Y. S. (2014). Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta Menggunakan Algoritma C4.5. *Jurnal Komuniti Volume VI, No.1*.
- Suhartono, Suparlan. 2009. Filsafat Pendidikan. Jogjakarta: Ar-Ruzz Media
- Sunarrihantono, Bimo. 2002. PHP Dan MySQL Untuk Web. Yogyakarta: Andi.
- Wahyudi, D. (2003). *Membuat Situs Menggunakan PHP Website*. Jakarta: PT Elex Media Komputindo.