

# Penerapan Algoritma C4.5 Untuk Penerimaan Siswa Baru Pada SMA Al-Azhar Medan

**Dedek Indra Gunawan Hts**

Universitas Potensi Utama

Jl. K. L. Yos Sudarso Km. 6,5 No. 3 A – Medan, 20241, Indonesia

dedek.indra@gmail.com

## **Abstract**

*Schools with good reputation are very concerned about the admission selection process. Al-Azhar Medan High School is very concerned about this admission process. Every year many parents register their children to enter Al-Azhar High School Medan. Besides having good school facilities the teachers also have good competencies. The number of students who register makes the selection process take a long time. The C4.5 algorithm is one algorithm that changes the fact that is quite large into a decision tree. This algorithm is expected to be able to solve problems that occur in Al-Azhar High School Medan.*

**Keywords:** *Decision Support System, C4.5 Algorithm, Acceptance of Students*

## **Abstrak**

*Sekolah yang bereputasi baik sangat memperhatikan proses seleksi penerimaan siswa/i nya. SMA Al-Azhar Medan sangat memperhatikan proses penerimaan ini. Setiap tahun banyak orangtua yang mendaftarkan anak-anaknya untuk masuk SMA Al-Azhar Medan. Selain memiliki fasilitas sekolah yang baik guru-gurunya juga memiliki kompetensi yang baik. Banyaknya siswa/i yang mendaftar menjadikan proses seleksi memakan waktu yang cukup lama. Algoritma C4.5 merupakan salah satu algoritma yang mengubah fakta yang cukup besar menjadi sebuah pohon keputusan. Dengan algoritma ini diharapkan mampu menyelesaikan permasalahan yang terjadi di SMA Al-Azhar Medan.*

**Kata kunci:** *Sistem Pendukung Keputusan, Algoritma C4.5, Penerimaan Siswa*

## **1. PENDAHULUAN**

SMA AL-Azhar adalah salah satu sekolah yang terbaik di Kota Medan. Setiap tahunnya banyak para orangtua siswa yang ingin mendaftarkan anak-anaknya ke sekolah tersebut. Banyaknya siswa yang mendaftar mengharuskan pihak sekolah mengadakan proses seleksi yang sangat ketat. Seleksi dilakukan agar pihak sekolah dapat mendidik siswa/i dengan maksimal. Para siswa harus mengikuti ujian Matematika, Bahasa Inggris dan Psikotes. Untuk masuk kelas Plus nilai ujian siswa harus di atas 6. Setelah proses tes kemampuan selesai, selanjutnya adalah para siswa harus mengikuti tes wawancara. Setelah tes wawancara selesai baru kemudian para siswa mengetahui hasilnya diterima di kelas Plus.

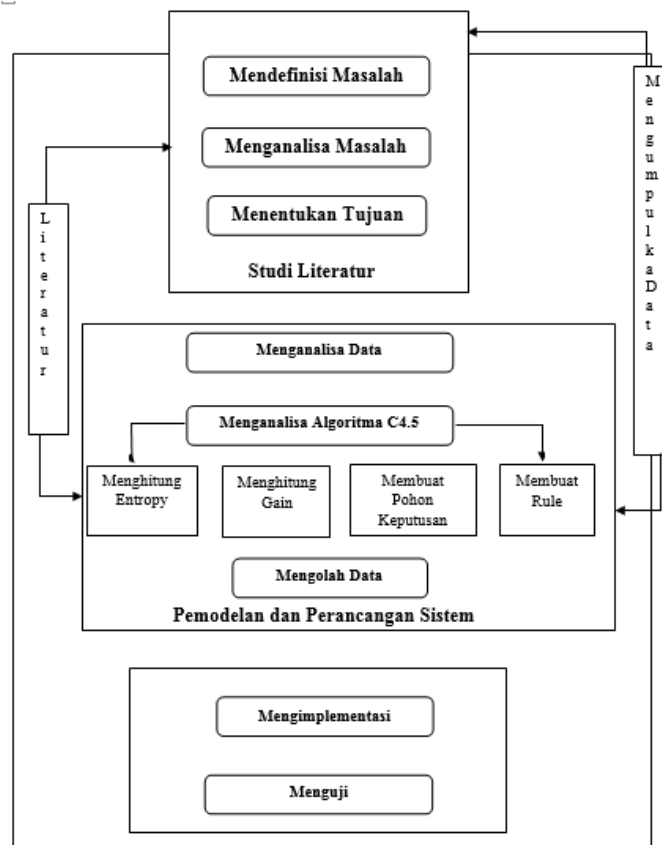
Proses penerimaan siswa di SMA Al-Azhar masih menggunakan perhitungan manual. Masing-masing tes dihitung baru kemudian diakumulasi. Hasil akumulasi tersebut yang menentukan apakah siswa tersebut masuk kelas Plus atau Akselerasi. Banyaknya siswa yang mengikuti ujian seleksi tentu membuat proses penilaian menjadi semakin lambat. Untuk

itulah dibutuhkan suatu metode untuk mempercepat proses penerimaan siswa baru di sekolah SMA Al-Azhar Medan.

Algoritma C4.5 merupakan salah satu metode yang bisa mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan[1]. Hasil tes seluruh setiap siswa diproses menggunakan algoritma C4.5 sehingga menghasilkan kesimpulan siapa saja siswa yang masuk kelas Plus atau Akselerasi. Dengan adanya sistem klasifikasi ini maka pihak sekolah dapat menentukan siapa saja siswa yang masuk kelas Plus dan Akselerasi. Sehingga proses penerimaan siswa menjadi lebih cepat.

## 2. METODOLOGI PENELITIAN

Metode penelitian sangat penting untuk membantu penulis dalam menyelesaikan masalah diatas. Metode penelitian dapat dilihat pada gambar kerangka kerja dibawah ini:



**Gambar 1.** Metode Penelitian

### a) Mendefinisi Masalah

Pada tahap ini dilakukan peninjauan ke sistem yang akan diteliti untuk mengamati serta melakukan eksplorasi lebih dalam dan menggali permasalahan yang ada pada sistem yang berjalan saat ini. Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian.

- b) Menganalisa Masalah  
Permasalahan yang ditemukan kemudian akan di analisa. Langkah dalam proses analisa masalah adalah langkah untuk memahami masalah yang telah ditentukan. Dengan menganalisa permasalahan yang telah ditentukan tersebut, maka diharapkan masalah tersebut dapat dipahami dengan baik.
- c) Menentukan Tujuan  
Berdasarkan pemahaman dari permasalahan yang telah di analisa, langkah berikutnya adalah menentukan tujuan yang akan dicapai dalam penelitian ini. Tujuan yang akan dicapai merupakan pengetahuan yang dapat digunakan untuk pengambilan keputusan dalam penerimaan siswa baru pada SMA AL-Azhar Medan.
- d) Mempelajari Literatur  
Penelitian ini dilakukan untuk melengkapi perbendaharaan kaidah, konsep, teori-teori yang mendukung dalam penyelesaian masalah dalam penelitian ini. Penelitian juga dilakukan melalui jurnal-jurnal yang ada hubungannya dengan penelitian maupun refrensi yang lain. Penelitian ini bertujuan untuk mengumpulkan data, baik data pokok maupun data pendukung, dimana semua data tersebut sangat dibutuhkan dalam penelitian.
- e) Mengumpulkan Data  
Mengumpulkan data untuk pelatihan dan pengujian *C4.5*. Semakin banyak data diperoleh, semakin baik dalam meyelesaikan masalahnya. Mengumpulkan data yang akurat dan membagi data tersebut ke dalam kriteria yang sudah ditentukan. Pembagian kriteria digunakan untuk mempermudah dalam pengelompokan data
- f) Menganalisa Algoritma *C4.5*  
Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada Algoritma *C4.5*, dimana tahapan yang dilakukan pada Algoritma *C4.5* dimulai dengan menghitung *entropy*, menghitung *gain*, membuat pohon keputusan dan membuat *rule*.
- g) Mengolahan Data  
Setelah proses analisa metode pada tahap ini dilakukan pengolahan data yang terdiri dari beberapa tahapan untuk menghasilkan pohon keputusan yaitu Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada.
- h) Mengimplementasi  
Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools.
- i) Menguji  
Pada tahap ini, dilakukan penilaian apakah perangkat lunak yang dikembangkan telah sesuai dengan tujuan yang diharapkan. Berikut ini adalah mekanisme pengujian yang dilakukan:

1. Membangun suatu kasus uji yaitu sekumpulan data atau situasi yang akan digunakan dalam pengujian.
2. Menentukan hasil yang akan diharapkan dengan cara melakukan proses perhitungan manual.
3. Menjalankan kasus pengujian

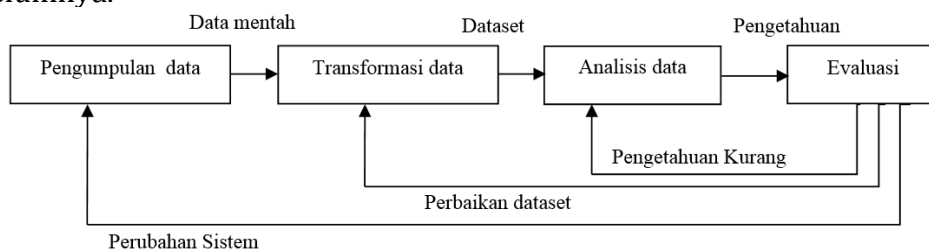
Melakukan perbandingan hasil pengujian dan hasil yang diharapkan, jika terdapat perbedaan hasil maka akan dilakukan perbaikan sesuai dengan kesalahan yang ditemukan.

### 2.1. Knowledge Discovery in Database

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar[2]. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

- a. *Data Selection*
- b. *Pre- processing / Cleaning*
- c. *Transformation*
- d. *Data mining*
- e. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

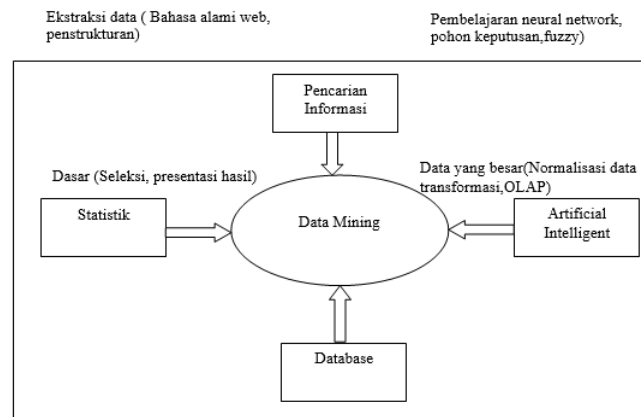


**Gambar 2.** Aliran Informasi dalam *Data Mining*

### 2.2. Data Mining

*Data mining* (DM) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database [3]. Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[4][5]. Hubungan yang dicari dalam *data mining* dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya

dalam dimensi produk kita dapat melihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungan juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek.



**Gambar 3.** Pengelompokan *Data Mining*

### 2.3. Klasifikasi

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bias digunakan untuk memprediksi kelas dari obyek yang label kelasnya tidak diketahui. Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning (fase training)*, dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi

### 2.4. Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain *ID3*, *CART*, dan *C4.5* [2]. Algoritma *C4.5* merupakan pengembangan dari algoritma *ID3*. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksikan aturan. Aturan dapat dengan mudah dipahami dengan alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari record pada kategori tertentu [2]. Secara umum algoritma *C4.5* untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus di bawah ini :

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

Di mana:

- $S$  = Himpunan kasus
- $A$  = Atribut
- $n$  = Jumlah partisi atribut A
- $|S_i|$  = Jumlah kasus pada partisi ke-i
- $|S|$  = Jumlah kasus dalam S

Sementara itu, perhitungan nilai *entropy* adalah seperti persamaan 2 di bawah ini :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Di mana:

- $S$  = Himpunan kasus
- $n$  = Jumlah partisi S
- $A$  = Fitur
- $P_i$  = Proporsi dari  $|S_i|$  terhadap S

Dalam hal ini mengubah *tree* yang dihasilkan dalam beberapa *rule* dimana jumlah *rule* sama dengan jumlah *path* yang dibangun dari *root* sampai *leaf node*. Ada istilah *Tree Pruning* yang dilakukan untuk menyederhanakan *tree* sehingga akurasi dapat bertambah. *Pruning* ada dua pendekatan, yaitu:

a. *Pre-pruning*

yaitu menghentikan pembangunan suatu *subtree* lebih awal (yaitu dengan memutuskan untuk tidak lebih jauh mempartisi data training). Saat seketika berhenti, maka *node* berubah menjadi *leaf (node akhir)*. *Node* akhir ini menjadi kelas yang paling sering muncul di antarsubset sampel.

b. *Post-pruning*

yaitu menyederhanakan *tree* dengan cara membuang beberapa cabang *subtree* setelah *tree* selesai dibangun. *Node* yang jarang dipotong akan menjadi *leaf (node akhir)* dengan kelas yang paling sering muncul.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Proses Pembentukan Pohon Keputusan

Berikut ini beberapa langkah dalam pembentukan pohon keputusan menggunakan algoritma C4.5

##### 3.1.1. Inisiasi

Berdasarkan tabel 1 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Wawancara, yaitu Sebesar 1,0695692. Dengan demikian,

*Wawancara* dapat menjadi node akar. Ada dua nilai atribut yaitu  $<7$  dan  $>7$ . Hasil perhitungan *entropy* dan *gain* diperoleh hasil seperti pada Tabel 1.

**Tabel 1.** Hasil Perhitungan Entropy dan Gain Akar

	Atribut	Isi/ Nilai	Jumlah Kasus	Hasil/ output		Entropy	Gain
				Diterima	Tidak Diterima		
Akar	Total		200				
	Nilai Tes Bahasa Inggris	$<5$	88	0	88	0	0,1089556
		5-7	40	10	30	0,0217293	
		$>7$	72	72	0	0,2905827	
	Nilai Tes Matematika	$<5$	120	0	120	0	0,0469234
		5-7	48	25	23	0,1090736	
		$>7$	32	32	0	0,129661	
	Nilai Tes Psikologi	$<5$	98	0	98	0	0,0604603
		5-7	54	23	31	0,0680559	
		$>7$	48	48	0	0,175355	
	Penghasilan orangtua	Menengah Keatas	172	172	0	0,835519	0,7185464
		Menengah Kebawah	28	0	28	0	
	Wawancara	$<7$	44	0	44	0	1,0695692
		$>7$	156	156	0	1,3712426	

### 3.1.2. Inisiasi Pertama

Berdasarkan tabel 2 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Penghasilan Orangtua, yaitu Sebesar 0,646867377. Dengan demikian, Penghasilan Orangtua menjadi cabang berikutnya. Ada dua nilai atribut yaitu Menengah Keatas dan Menengah Kebawah. Hasil perhitungan *entropy* dan *gain* diperoleh hasil seperti pada Tabel 2.

**Tabel 2.** Hasil Perhitungan Entropy dan Gain Level I

	Atribut	Isi/ Nilai	Jumlah Kasus	Hasil/ output		Entropy	Gain
				Diterima	Tidak Diterima		
Level I	Wawancara		156				
	Nilai Tes Bahasa Inggris	$<5$	88	0	88	0	0,033170831
		5-7	34	0	34	0	
		$>7$	34	34	0	0,152195578	
	Nilai Tes Matematika	$<5$	70	0	120	0	0,147097054
		5-7	30	0	30	0	
		$>7$	56	56	0	0,409770366	
	Nilai Tes Psikologi	$<5$	98	0	98	0	0,042741749
		5-7	20	0	20	0	
		$>7$	38	38	0	0,175466128	
	Penghasilan orangtua	Menengah Keatas	128	128	0	0,788369615	0,646867377
Menengah Kebawah		28	0	28	0		

### 3.1.3. Inisiasi Kedua

Berdasarkan tabel 3 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Nilai Psikotes, yaitu Sebesar 0,093781188. Dengan demikian,

Nilai Psikotes menjadi cabang berikutnya. Ada tiga nilai atribut yaitu <5, 5-7, >7. Hasil perhitungan *entropy* dan *gain* diperoleh hasil seperti pada Tabel 3.

**Tabel 3.** Hasil Perhitungan Entropy dan Gain Level II

	Atribut	Isi/ Nilai	Jumlah Kasus	Hasil/ output		Entropy	Gain
				Diterima	Tidak Diterima		
Level II	Wawancara, Penghasilan Orangtua		128				
	Nilai Tes Bahasa Inggris	<5	85	0	88	0	0,025307008
		5-7	18	0	18	0	
		>7	25	25	0	0,12957188	
	Nilai Tes Matematika	<5	77	0	77	0	0,053918258
		5-7	25	10	15	0,062237482	
		>7	26	26	0	0,205599997	
	Nilai Tes Psikologi	<5	70	0	70	0	0,093781188
		5-7	12	6	6	0,017310017	
>7		44	44	0	0,268097088		

### 3.1.4. Inisiasi Ketiga

Berdasarkan tabel 4 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Nilai Matematika, yaitu Sebesar 0,664037053. Dengan demikian, Nilai Psikotes menjadi cabang berikutnya. Ada tiga nilai atribut yaitu <5, 5-7, >7. Hasil perhitungan *entropy* dan *gain* diperoleh hasil seperti pada Tabel 4.

**Tabel 4.** Hasil Perhitungan Entropy dan Gain Level III

	Atribut	Isi/ Nilai	Jumlah Kasus	Hasil/ output		Entropy	Gain
				Diterima	Tidak Diterima		
Level III	Wawancara, Penghasilan Orangtua, Nilai Tes Psikologi		50				
	Nilai Tes Bahasa Inggris	<4	0	0	0	0	0,44532472
		5-7	15	0	15	0	
		>7	35	35	0	0,636178172	
	Nilai Tes Matematika	<5	0	0	120	0	0,664037053
		5-7	10	0	10	0	
>7		40	40	0	0,830046317		

### 3.1.5. Inisiasi Keempat

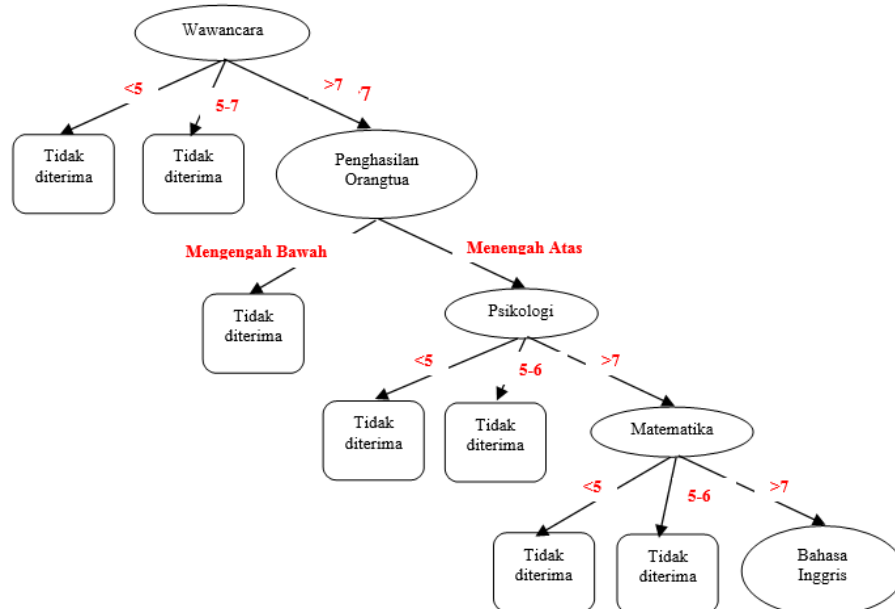
Berdasarkan tabel 5 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Nilai Matematika, yaitu Sebesar 0,340855098. Dengan demikian, Nilai Psikotes menjadi cabang berikutnya. Ada tiga nilai atribut yaitu <5, 5-7, >7. Hasil perhitungan *entropy* dan *gain* diperoleh hasil seperti pada Tabel 5.



**Tabel 5.** Hasil Perhitungan Entropy dan Gain Level IV

Level IV	Atribut	Isi/ Nilai	Jumlah Kasus	Hasil/ output		Entropy	Gain
				Diterima	Tidak Diterima		
Level IV	Wawancara, Penghasilan Orangtua, Nilai Tes Psikologi, Nilai Tes Matematika		40				0,340855098
	Nilai Tes Bahasa Inggris	<4	0	0	88	0	
		5-7	15	0	40	0	
>7		25	25	0	0,545368157		

Hasil perhitungan *Entropy* dan *Gain* Mencari Cabang Level 1 dalam bentuk pohon keputusan dapat dilihat pada gambar 4.



**Gambar 4.** Pohon Keputusan Menggunakan Algoritma C4.5

#### 4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

- Model pohon keputusan yang dihasilkan adalah menjadikan Wawancara sebagai akar, Kemudian Penghasilan Orangtua sebagai cabang pada Level 1, Psikologi menjadi cabang pada Level 2, dan Matematika menjadi cabang pada Level 3 dan Bahasa Inggris menjadi cabang pada Level terakhir (Level 4).
- Proses perhitungan dan penyeleksian seleksi penerimaan calon siswa/i baru menjadi lebih mudah dan cepat dengan menggunakan sistem pendukung keputusan ini.

#### DAFTAR PUSTAKA

- [1] Adriyansyah, Dedy. "Faktor-faktor Apa Saja Yang Mempengaruhi Konsumen Listrik Bagi Rumah Tangga Masyarakat Kelurahan Tembung Kecamatan Percut Sei Tuan Kabupaten Deli Serdang." *Universitas Sumatera Utara, Medan* (2011).
- [2] Kusrini, (2009). *Algoritma Data Mining*, Andi Offcet, Yogyakarta
- [3] Arga, Dwi Asih Septa, Edhy Sutanta, and Uning Lestari. "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritme C4. 5 (Studi Kasus: Jurusan Teknik Informatika, Institut Sains & Teknologi Akprind Yogyakarta)." *Jurnal Script* 5.1 (2017).
- [4] Widayu, Hikma, et al. "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4. 5." *Media Informatika Budidarma* 1.2 (2017).
- [5] Harahap, Fitriana. "Penerapan data Mining dalam Pemilihan Mobil menggunakan Algoritma C4. 5." *Voice Of Informatics* 7.1 (2018).