

# Data Mining Penentuan Tenaga Perawat Di RSUD Sultan Sulaiman Dengan Algoritma C4.5

Wenika Hidayati

STMIK Pelita Nusantara, Jl. Iskandar Muda No.1 Medan, Sumatera Utara, Indonesia 20154

Email : [wenikahidayati49@gmail.com](mailto:wenikahidayati49@gmail.com)

INFORMASI ARTIKEL	ABSTRAK
<b>Kata Kunci:</b> Penentuan Data Mining	Rumah sakit adalah sebuah instansi yang bergerak dibidang pelayanan kesehatan yang didalamnya terdapat beberapa profesi khusus yang dapat melakukan pelayanan kesehatan kepada masyarakat, yakni dokter, Bidan dan Perawat dan profess lainnya. Dalam pembahasan ini, masalah yang timbul dan dapat diangkat menjadi studi kasus untuk mengetahui hasil dan informasi dari setiap proses yang dilakukan pada data mining dengan algoritma c4.5 yaitu tenaga perawat. Namun terdapat beberapa kendala untuk menentukan tenaga perawat yang akan dinyatakan lulus atau gagal dan diterima untuk bekerja serta dapat melakukan pelayanan kesehatan kepada masyarakat khususnya pasien yang datang untuk berobat. Oleh karena itu diperlukan metode untuk mengidentifikasi tenaga perawat pada sebuah Rumah Sakit. Data Mining dengan Algoritma c4.5 dapat digunakan untuk melakukan prediksi atau klasifikasi terhadap tenaga perawat yang layak untuk melakukan pelayanan kesehatan di rumah sakit dengan cara membuat pohon keputusan berdasarkan data yang sudah ada. Penelitian ini bertujuan untuk menerapkan data mining algoritma C4.5 dalam menentukan tenaga perawat berdasarkan 4 atribut yang digunakan yakni Akreditasi, IPK, Umur, Nilai yang masing-masing kriterianya sudah ditentukan terlebih dahulu. Hasil penelitian berupa pohon keputusan yang didapat dari proses data mining dengan algoritma C4.5 akan memberikan informasi penentuan tenaga perawat di RSUD Sultan Sulaiman.
<b>Keywords:</b> Determination Data Mining Algoritma C4.5	<b>ABSTRACT</b> The hospital is an agency engaged in health services in which there are a number of special professions that can provide health services to the community, namely doctors, midwives and nurses and other professes. In this discussion, problems that arise and can be raised into case studies to find out the results and information of each process carried out in data mining with the c4.5 algorithm, namely nurses. However, there are several obstacles to determine the nurses who will be declared passed or failed and accepted to work and can provide health services to the community, especially patients who come for treatment. Therefore we need a method to identify nurses in a hospital. Data Mining with c4.5 Algorithm can be used to make predictions or classifications of nurses who are eligible to perform health services in hospitals by making decision trees based on existing data. This study aims to apply the data mining algorithm C4.5 in determining nurses based on 4 attributes used, namely Accreditation, GPA, Age, and the value of each criterion has been determined in advance. The results of the study in the form of a decision tree obtained from the data mining process with the C4.5 algorithm will provide information on the determination of nurses in the Sultan Sulaiman Regional Hospital.

## I. Pendahuluan

Data mining merupakan serangkaian proses untuk mendapatkan informasi berbasis komputer yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak teknin dalam pengerjaannya, diantaranya yaitu Algoritma C4.5, *Naive*

Bayes dan lain sebagainya. Data mining ini dapat digunakan sebagai penentuan tenaga perawat berdasarkan kriteria-kriteria yang telah ditentukan. Cara kerja data mining ini mencakup seluruh tahap pengambilan masalah, memilih data yang relevan dan menentukan pemodelan yang digunakan dalam proses penyelesaian masalah, pemecahan dan solusi masalah.

RSUD Sultan Sulaiman merupakan rumah sakit yang terletak sangat strategis di Jl. Negara Km.58 Sei Rampah. Rumah sakit ini merupakan rumah sakit kelas C yang mengembangkan pelayanan kesehatan secara profesional kepada masyarakat. Di RSUD Sultan Sulaiman, jumlah dan jenis tenaga keperawatan berdasarkan jenjang pendidikan yaitu, D3 Keperawatan, S1 Keperawatan, dan Keperawatan berprofesi Ners yang tersebar di unit kerja rawat jalan dan rawat inap, sedangkan syarat untuk bekerja dan melakukan pelayanan kesehatan di RSUD Sultan Sulaiman bukan hanya berdasarkan dengan pendidikan, namun juga dapat ditentukan dari aspek lainnya seperti, Umur, Akreditasi, IPK dan hasil ujian yang sesuai dengan kriteria dan pimpinan rumah sakit. Menurut data seksi bagian kepegawaian dan seksi keperawatan, saat ini di RSUD Sultan Sulaiman menggunakan data manual untuk menentukan lulus atau tidaknya tenaga keperawatan yang bekerja dan melakukan pelayanan kesehatan di RSUD Sultan Sulaiman.

Berdasarkan penjelasan dari latar belakang pada penelitian ini, maka identifikasi masalah adalah sebagai berikut :

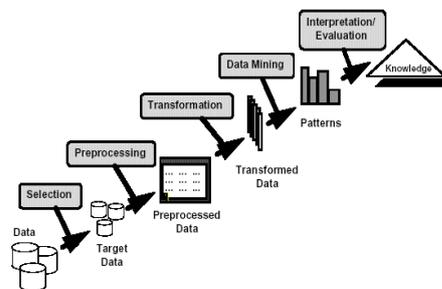
1. Penentuan tenaga perawat masih dilakukan secara manual baik berupa laporan.
2. Pengolahan data kriteria dalam penentuan tenaga perawat menggunakan sistem manual sehingga proses seleksi dan penentuan memakan waktu yang cukup lama.
3. Penyimpanan data perawat dan penentuan tenaga perawat tidak tersusun dengan rapi dan aman.

## II. Metode

### 1. Metode Penelitian

Metode Penelitian adalah rangkaian cara terstruktur atau sistematis yang digunakan oleh para peneliti dengan tujuan mendapatkan jawaban yang tepat atas apa yang menjadi pertanyaan pada objek penelitian. Metode ini menjelaskan tentang metode apa yang digunakan dalam penelitian yang dilakukan, maka digunakan metodologi. Metodologi adalah ilmu tentang kerangka kerja untuk melaksanakan penelitian yang bersistem mengenai suatu cara atau metode, cabang ilmu logika yang berkaitan dengan prinsip umum pembentukan pengetahuan (*Knowledge*).

Data Mining merupakan salah satu dari rangkaian *Knowledge Discovery in Database (KDD)*. KDD berhubungan dengan Teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Data Mining dapat dibagi menjadi beberapa tahap seperti terlihat pada gambar dibawah ini :



Gambar 1. Proses KDD

### Algoritma C4.5

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

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

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. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk decision tree berdasarkan kriteria-kriteria pembentuk decision. Rumus algoritma C4.5 terbagi menjadi 2 rumus, yang pertama rumus untuk mencari Gain

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain*, seharusnya kita harus mencari nilai *entropy* terlebih dahulu. Rumusnya seperti persamaan berikut.

$$Entropy(S) - (x + a)^n = \sum_{j=1}^k -P_j \text{Log}_2 P_j$$

Keterangan :

- S : Himpunan kasus
- k : Jumlah partisi S
- p<sub>j</sub> : Probabilitas yang didapat dari jumlah (Ya/Tidak) dibagi total kasus

Sementara itu, penghitungan nilai *gain* dapat dilihat pada persamaan 2 berikut.

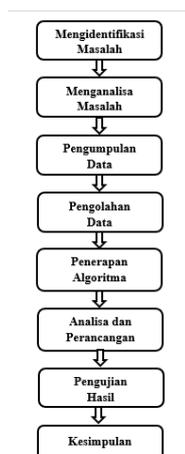
$$Gain(S, A) = Entropy(S) - (x + a)^n = \sum_{i=1}^n -\frac{S_i}{S} * Entropy(S_i)$$

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

### Kerangka Kerja Penelitian

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya susunan kerangka kerja (Frame Work) yang jelas pada tahapan-tahapannya. Kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Berikut langkah-langkah yang dilakukan sesuai dengan alur rancangan penelitian penulis sebagai berikut



Gambar 2. Kerangka Kerja Penelitian

### III. Hasil dan Pembahasan

Analisa sistem merupakan tahap awal dalam perancangan dan pengembangan sebuah sistem yang akan dirancang, karena tahap inilah akan diukur dan dievaluasi tentang kinerja dari sistem yang dirancang. Identifikasi terhadap masalah-masalah yang ada dan langkah-langkah untuk kebutuhan perancangan yang diharapkan.

Tabel 1. Hasil dari *Cleaning Data* Tenaga Perawat

No	Nama Perawat	Akreditasi	Umur	IPK	Nilai
1	Afriani Yulma	26	3,19	B	83
2	Mariska Nurhabibi	22	3.63	B	92
3	Nurul Bariah	25	3.01	B	79
4	Halim Fadila	23	3.69	B	85
5	Aprilin Syahputra Lubis	26	3,42	B	84

6	Mariska Nurhabibi	22	3,63	B	92
7	Dedek Kurniawan	21	3,02	A	78
8	Asiah	25	3,46	A	90
9	Ayu Annisa	24	3,41	B	80

Dari Proses di atas dihasilkan tabel yang siap untuk proses Klasifikasi, dengan atribut dan nilai atribut yang ada pada tabel 2 berikut tabel atribut yang didapat.

Tabel 2 Atribut perawat

No	Atribut	Keterangan
1	Akreditasi	Merupakan atribut dari akreditasi kampus yaitu "A" dan "B".
2	Umur	Merupakan bukti tertulis yang dapat dilihat dari KTP
3	IPK	Merupakan nilai tertulis yang diberikan sebuah fakultas / universitas yang dapat dilihat dari ijazah atau transkrip nilai.
4	Nilai	Merupakan Nilai yang diperoleh dari hasil ujian masing-masing peserta

Dari daftar item yang ada diatas, maka data yang di gunakan yaitu akreditasi A dan B, umur 21 s/d 28 tahun, IPK 2,50 s/d 4,00, Nilai >55 s/d 100 untuk itu diambil beberapa data untuk dijadikan sampel pada penentuan tenaga perawat di RSUD Sultan Sulaiman Kab. Serdang Bedagai. Berikut daftar nama-nama perawat yang dijadikan sampel di RSUD Sultan Sulaiman

Menghitung jumlah kasus untuk keputusan lulus atau gagal, jumlah kasus untuk keputusan lulus atau gagal, Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut jumlah tenaga perawat, perawat yang layak untuk diterima, dan perawat yang tidak layak untuk di terima. Setelah itu, lakukan penghitungan gain untuk setiap atribut. Hasil perhitungan ditunjukkan oleh tabel 3.

Tabel 3. Perhitungan C4.5 (1)

Nama Variabel	Keterangan	Sum (Nilai)	Sum (Ya)	Sum (Tidak)	Entropi	Gain
Akreditasi	A	2	2	0	0	0,225
	B	7	3	4	0,985	
Umur	<= 25	6	5	1	0,65	0,558
	>= 25	3	0	3	0	
IPK	<= 3,50	8	4	4	1	0,102
	>= 3,51	1	1	0	0	
Nilai	< 75	1	0	1	0	0,954
	> 75	8	5	3	0,954	

Entropi :

$$1. \text{ Akreditasi A} = (-2/2) \times \text{Log}_2(2/2) + (-0/2) \times \text{Log}_2(0/2) = 0$$

$$\text{ Akreditasi B} = (-3/7) \times \text{Log}_2(3/7) + (-4/7) \times \text{Log}_2(4/7) = 0,985$$

Gain untuk akreditasi :

$$= 0,99107 - ((2/9) * 0) + ((7/9) * 0,985)$$

$$= 0,225$$

Entropi :

$$2. \text{ Umur } <=25 = (-5/6) \times \text{Log}_2(5/6) + (-1/6) \times \text{Log}_2(1/6) = 0,65$$

$$\text{ Umur } >=25 = (-0/3) \times \text{Log}_2(0/3) + (-3/3) \times \text{Log}_2(3/3) = 0$$

Gain untuk akreditasi :

$$= 0,99107 - (6/9) * 0,65 + ((3/9) * 0)$$

$$= 0,558$$

Entropi :

$$3. \text{ IPK } <= 3,50 = (-4/8) \times \text{Log}_2(4/8) + (-4/8) \times \text{Log}_2(4/8) = 1$$

$$\text{ IPK } >= 3,51 = (-1/1) \times \text{Log}_2(1/1) + (-0/1) \times \text{Log}_2(0/1) = 0$$

Gain untuk akreditasi :

$$= 0,99107 - (4/9) * 1 + ((1/9) * 0)$$

$$= 0,102$$

Entropi :

$$4. \text{ Nilai } <= 75 = (-0/1) \times \text{Log}_2(0/1) + (-1/1) \times \text{Log}_2(1/1) = 0$$

$$\text{ Nilai } >= 75 = (-5/8) \times \text{Log}_2(5/8) + (-3/8) \times \text{Log}_2(3/8) = 0,954$$

Gain untuk akreditasi :  
 $= 0,99107 - (4/10) * 0,81127812445) + ((6/10) * 0)$   
 $= 0,99107 - 0,32451124978$   
 $= 0,143$

Berdasarkan hasil diketahui bahwa atribut dengan Gain tertinggi adalah Umur, yaitu sebesar 0,558 Dengan demikian Umur dapat menjadi node akar, nilai atribut sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya LULUS, sehingga tidak perlu dilakukan perhitungan lebih lanjut.

Menghitung jumlah kasus, jumlah kasus untuk keputusan LULUS, jumlah kasus untuk keputusan GAGAL, dan entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut Akreditasi, Nilai, IPK yang dapat menjadi node akar dari nilai atribut. Setelah itu lakukan perhitungan Gain untuk masing-masing atribut. Hasil tersebut ditunjukkan oleh tabel berikut:

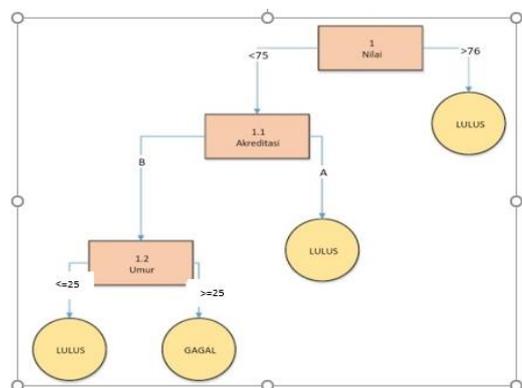
Tabel 4. Perhitungan C4.5 (2)

Nama Variabel	Keterangan	Sum (Nilai)	Sum (Ya)	Sum (Tidak)	Entropi	Gain
Akreditasi	A	2	2	0	0	
	B	4	3	1	0,811	0,109
Nilai	$\leq 75$	0	0	0	0	
	$\geq 75$	6	5	1	0,65	0
IPK	$\leq 3.50$	5	4	1	0,722	
	$\geq 3.51$	1	1	0	0	0,048

- Akreditasi A =  $(-(2/2) \times \text{Log}_2 (2/2) + (-(0/2) \times \text{Log}_2 (0/2)) = 0$   
 Akreditasi B =  $(-(3/4) \times \text{Log}_2 (3/4) + (-(1/4) \times \text{Log}_2 (1/4)) = 0,811$   
 Gain untuk akreditasi :  
 $= 0,99107 - ((2/4) * 0) + ((2/4) * 0,811)$   
 $= 0,109$
- Nilai  $\leq 75 = (-(0/0) \times \text{Log}_2 (0/0) + (-(0/0) \times \text{Log}_2 (0/0)) = 0$   
 Nilai  $\geq 75 = (-(5/6) \times \text{Log}_2 (5/6) + (-(1/6) \times \text{Log}_2 (1/6)) = 0,65$   
 Gain untuk akreditasi :  
 $= 0,99107 - (0/6) * 0) + ((0/6) * 0)$   
 $= 0$
- IPK  $\leq 3,50 = (-(4/5) \times \text{Log}_2 (4/5) + (-(1/5) \times \text{Log}_2 (1/5)) = 0,722$   
 IPK  $\geq 3,51 = (-(1/1) \times \text{Log}_2 (1/1) + (-(0/1) \times \text{Log}_2 (0/1)) = 0$   
 Gain untuk akreditasi :  
 $= 0,99107 - (5/1) * 0,722) + ((5/1) * 0)$   
 $= 0,048$

Berdasarkan hasil diketahui bahwa atribut dengan Gain tertinggi adalah Akreditasi, yaitu sebesar 0,109. Dengan demikian Akreditasi dapat menjadi node cabang dari nilai atribut. Ada 2 nilai atribut dari Akreditasi yaitu A dan B. Dari ke

2 nilai atribut tersebut, nilai A sudah mengklasifikasikan kasus menjadi LULUS dan nilai atribut B sudah mengklasifikasikan kasus menjadi GAGAL, kemudian dari pada itu tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut B masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat di gambarkan pohon keputusan sementara seperti gambar berikut :



Gambar 3. Pohon Keputusan

1. Jika Umur <25, Akreditasi B, IPK >3,51 = Lulus
2. Jika Umur <25, Akreditasi B, IPK >3,50, Nilai <85 = Lulus
3. Jika Umur <25, Akreditasi B, IPK >3,50, Nilai >85 = Lulus
4. Jika Umur <25, Akreditasi A = Lulus
5. Jika Umur >25 = Gagal

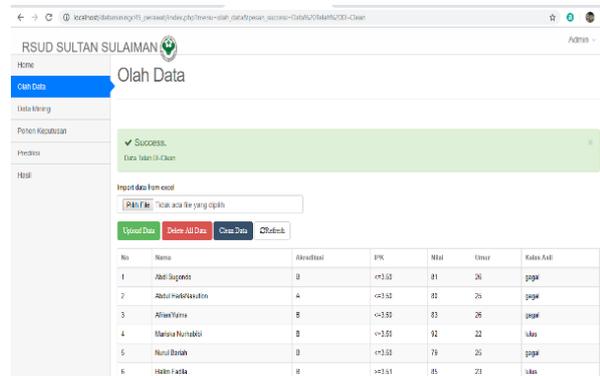
Menghitung jumlah kasus, jumlah kasus untuk keputusan LULUS, jumlah kasus untuk keputusan GAGAL, dan entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut, dan Status yang dapat menjadi node akar dari nilai atribut. Dengan memperhatikan pohon keputusan pada gambar sebelumnya, diketahui bahwa kasus sudah masuk dalam kelas. Dengan demikian pohon keputusan terakhir yang terbentuk.

### Pembahasan

Setelah melalui tahap perancangan sistem, tahap berikutnya adalah implementasi sistem. Adapun hasil implementasi dari analisis dan perancangan pada bab sebelumnya dapat dilihat seperti pada tampilan berikut ini.

#### 1. Form Olah Data

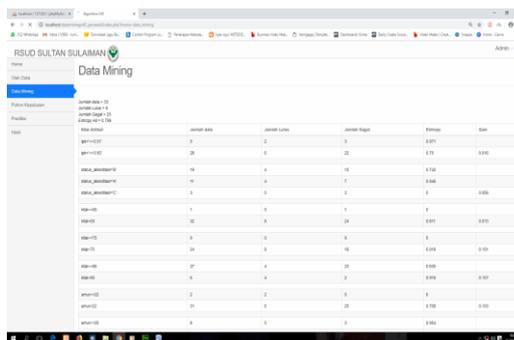
Halaman ini merupakan halaman untuk memasukkan data yang akan dilatih menggunakan algoritma C4.5. Pada halaman ini user dapat memasukkan data dengan cara mengupload data berekstensi \*.xls. Selain itu user dapat menghapus seluruh data dengan memilih tombol Hapus Data.



Gambar 4. Form Olah Data

#### 2. Form Proses Data Mining

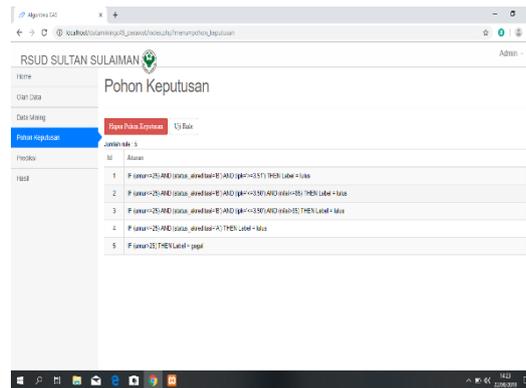
Pada halaman ini akan ditampilkan jumlah data yang dimining, jumlah data yang LULUS dan jumlah data yang GAGAL. Selain itu akan ditampilkan juga nilai Entropy dan nilai Gain.



Gambar 5. Form Proses Data Mining

#### 3. Form Pohon Keputusan

Setelah data diinputkan data akan ditransformasi dan diproses untuk menghasilkan pohon keputusan. Pohon keputusan akan terbentuk jika data yang diinputkan adalah data yang memang mempunyai pola tertentu dan mempunyai jumlah yang cukup besar. Pada halaman ini juga, pengguna dapat menguji pohon keputusan yang terbentuk dengan memilih tombol Uji Rule.



Gambar 6. Form Pohon Keputusan

#### IV. Kesimpulan

Dari hasil penelitian yang dilakukan penulis, maka diambil kesimpulan dan saran sebagai berikut :  
Dari hasil analisa, perancangan, dan implementasi dapat diperoleh kesimpulan sebagai berikut

1. Sistem yang dibangun dengan Algoritma C4.5 dapat membentuk pohon keputusan mulai dari akar sampai ke cabang terakhir yang terbentuk.
2. Sistem yang dibangun dapat digunakan untuk membantu mengklasifikasikan perawat yang memenuhi persyaratan berdasarkan kriteria-kriteria yang telah ditentukan.
3. Setelah menghitung menggunakan entropy dan gain dari data yang ada, Algoritma C4.5 telah dapat mengimplementasikan klasifikasi perawat berdasarkan kriteria penilaian
4. Sistem dapat menghitung jumlah kasus dengan keputusan LULUS dan GAGAL berdasarkan nilai entropy dari semua kasus berdasarkan atribut.

#### Daftar Pustaka

- [1] P. M. Hasugian, "Pengujian Algoritma Apriori Dengan Aplikasi Weka Dalam," *J. Mantik Penusa*, 2017.
- [2] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017.
- [3] A. S. Sunge, "Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 ( Studi Kasus : PT Hankook Tire Indonesia )," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 23–24, 2018.
- [4] G. F. Mandias, "Pemanfaatan Data Mining Untuk Prediksi Mahasiswa Peserta ' Capping Day ' di Universitas Klabat Data Mining for Student Prediction Participant of " Capping Day " at Universitas Klabat Using the C4 . 5 Algorithm," pp. 86–93.
- [5] K. Zaman, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4 . 5 UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN REHABILITASI SOSIAL RUMAH TIDAK LAYAK HUNI ( STUDI KASUS DI PEMERINTAHAN KABUPATEN SOLOK SELATAN )," vol. 3, no. 2, pp. 12–24, 2016.
- [6] G. Syahputra, "Penerapan Algoritma C4 . 5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai ( Studi Kasus : PT . Multi Pratama Nauli Medan )," *J. Mantik Penusa*, vol. 16, no. 2, 2014.
- [7] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4 . 5 Sebagai Dasar Pemberian Kredit," 2016.
- [8] R. V Palit, Y. D. Y. Rindengan, and A. S. M. Lumenta, "Rancangan Sistem Informasi Keuangan Gereja Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang," vol. 4, no. 7, pp. 1–7, 2015.