

# Memprediksi Tingkat Kelulusan Peserta Kursus Per-tahun dengan Algoritma Data Mining C4.5 dan RapidMiner

Iswati Nur Kholifah<sup>1</sup>

**Abstrak**— Lembaga Kursus adalah sebuah lembaga pendidikan yang berada di luar lingkup sekolah maupun pekerjaan, lembaga ini biasanya diselenggarakan oleh masyarakat dan pelajar ataupun mahasiswa. Banyaknya peserta yang mendaftar ke lembaga kursus, karena dapat membantu mereka dalam memperoleh wawasan maupun pengalaman dalam pekerjaan. Untuk mendukung minat masyarakat yang efektif, penulis menerapkan metode algoritma C4.5 untuk memberikan solusi dalam memprediksi tingkat kelulusan serta minat masyarakat yang diambil berdasarkan data pada tahun lalu. Pada tabel komputer, kesimpulan akhir data banyak peserta yang lulus tepat waktu sekitar 70% sedangkan yang tidak lulus tepat waktu hanya 30% menggunakan aplikasi rapid miner. Berdasarkan uji algoritma C45 dan aplikasi RapidMiner memiliki tingkat akurasi kurang dari 100%. Sehingga, dapat membantu dalam menentukan dan melihat seberapa minat masyarakat untuk kursus mengemudi serta kursus komputer sesuai dengan kebutuhan dan keinginan.

**Kata Kunci:** Algoritma C4.5, Data Mining, Pohon Keputusan

**Abstract**— Course Institute is an educational institution that is outside the scope of school and work, this institution is usually run by the community and students or students. The large number of participants who register to course institutions, because it can help them gain insight and experience in work. To support effective public interest, the authors apply the C4.5 algorithm method to provide solutions in predicting graduation rates and public interest based on data from last year. In the computer table, the final conclusion is that 70% of the participants passed on time while only 30% did not pass on time using the rapid miner application. Based on the C45 algorithm test and the RapidMiner application, it has an accuracy rate of less than 100% so that it can help determine and see how much public interest in driving courses and computer courses according to their needs and desires.

**Keywords:** Algorithm C4.5, Data Mining, Decision Tree

## I. PENDAHULUAN

Di zaman sekarang penting bagi seseorang meningkatkan ilmu pengetahuan agar bisa mengikuti perkembangan zaman yang semakin maju seperti saat ini,

terutama ilmu di bidang digital yang tentunya sangat diperlukan baik dari segi pendidikan maupun pekerjaan<sup>[15]</sup>. Dari kalangan masyarakat saja, saat ini masih banyak yang memiliki kekurangan dalam hal pengetahuan maupun pengalaman dalam pekerjaan, oleh sebab itu maka diperlukan adanya lembaga kursus untuk membantu dalam hal pengetahuan maupun pengalaman dalam pekerjaan.

Lembaga Kursus adalah sebuah lembaga pendidikan yang berada di luar lingkup sekolah maupun pekerjaan, lembaga ini biasanya diselenggarakan oleh masyarakat dan pelajar ataupun mahasiswa<sup>[7]</sup>. Lembaga kursus ini bertujuan untuk memberikan pengetahuan, lapangan pekerjaan, serta meningkatkan minat seseorang dalam keingintahuan akan ilmu. Kursusnya juga memiliki berbagai macam jenis, dari mulai kursus tentang akademik maupun non akademik, kursus tentang mengemudi berbagai kendaraan dan macam kursus lainnya.

Banyaknya peserta yang mendaftar ke lembaga kursus, karena dapat membantu mereka dalam mendapat wawasan maupun pengalaman dalam pekerjaan. Dikarenakan banyaknya peserta setiap tahunnya, membuat lembaga kursus perlu untuk menghitung data kelulusan per tahun, untuk dapat mengetahui daftar peminat dalam pendaftar kursus tersebut. Karena itu lembaga kursus perlu menerapkan data mining untuk menghitung daftar kelulusan tersebut<sup>[16]</sup>.

Algoritma C4.5 merupakan metode mining data yang direkomendasikan untuk menghitung daftar kelulusan pada lembaga kursus tersebut. Algoritma C4.5 yaitu hasil dari pengembangan algoritma ID3 (Iterative Dichotomiser) yang dikembangkan oleh J. Ross Quinlan pada tahun 1979<sup>[2],[14]</sup>. Algoritma C4.5 ini digunakan dalam sebuah klarifikasi yang menggunakan pohon keputusan. Algoritma ini memiliki kelebihan yaitu pada hasil yang bermodel pohon, memiliki tingkat akurasi yang dapat diterima sehingga memudahkan dalam pemahaman algoritmanya.

Tujuan dari penelitian ini yaitu untuk mengetahui hasil kelulusan setiap tahunnya pada lembaga kursus dengan menggunakan metode Algoritma C4.5. Pembuatan data mining dalam Algoritma yang akan membantu Lembaga kursus untuk mengetahui kelulusan setiap tahunnya<sup>[10],[18]</sup>.

<sup>1</sup> Mahasiswa, Jurusan Sistem Informatika Fakultas Ilmu Komputer Universitas Darwan Ali Sampit, Jln.Batu Berlian, Sampit Kalimantan Tengah, INDONESIA No. 10 Tlp. (0531) 33336, Fax. (0531)33342; e-mail: iswatinurkholifah999@gmail.com)

## II. METODE PENELITIAN

Metode yang akan digunakan dalam penulisan jurnal ini yaitu metode kuantitatif karena pada dasarnya metode ini merupakan proses keilmuan untuk memperoleh data secara sistematis dan juga berdasarkan bukti, berikut metode untuk penelitian dalam jurnal ini [21], [5].

### A. Study Literatur

Metode ini adalah sebuah pengumpulan data melalui jurnal ilmiah, penelitian sejenis, buku, artikel dan sumber bacaan lainnya yang berkaitan dengan penelitian. Sehingga data atau informasi yang telah didapatkan dari Kursus tersebut dapat dijadikan sebagai rujukan untuk memperkuat landasan dan argumentasi-argumentasi yang telah ada.

Studi literatur adalah sebuah rangkaian kegiatan yang berkaitan dengan metode yang melakukan pengumpulan data membaca, pustaka, mencatat, dan juga mengolah bahan-bahan penelitian. Tujuan dari metode ini adalah untuk menunjukkan berbagai teori-teori yang sering ada dengan permasalahan yang sering terjadi dan sedang dihadapi dan diteliti sebagai bahan dalam pembahasan untuk hasil penelitian

Pengertian lain dari studi literatur adalah untuk mencari referensi teori yang sering terjadi dengan kasus atau permasalahan yang sering ditemukan. Referensi ini biasanya dapat dicari dari jurnal, artikel laporan, buku, dan juga situs-situs di web atau internet. Keluaran atau *output* dari *studi literatur* ini adalah akan membuat terkoleksinya referensi yang relevan dengan perumusan dalam masalah yang sedang di teliti.

Secara umum *studi literatur* adalah sebuah cara yang digunakan untuk menyelesaikan masalah atau persoalan dengan cara menelusuri sumber-sumber tulisan yang telah dibuat sebelumnya. *Studi literatur* ini sangat sering ditemukan dengan mencari di studi pustaka. Dalam penelitian yang akan dijalankan, tentunya seorang peneliti harus mempunyai wawasan yang banyak dan luas dalam hal objek yang akan diteliti [25]. Jika peneliti tidak mempunyai wawasan yang luas terkait dengan objek yang dibahas, maka dapat dipastikan peneliti tersebut akan mempunyai persentase yang memungkinkan peneliti tersebut akan mengalami kegagalan saat akan melakukan penelitian.

untuk membantu penelitian dalam mencari sebuah teori penjelasan seperti Algoritma C4.5 beserta teori tentang lembaga Kursus bagaimana masyarakat memperoleh penambahan ilmu pengetahuan.

### B. Algoritma C4.5

Algoritma *data mining* C4.5 adalah salah satu algoritma yang sering digunakan untuk melakukan sebuah klasifikasi atau segmentasi atau pengelompokan dan mempunyai sifat prediktif [19]. Klasifikasi adalah salah satu proses yang ada

pada *data mining* yang memiliki tujuan untuk menemukan pola pada data yang berharga yang memiliki ukuran relatif besar hingga mempunyai data yang sangat besar [22].

Algoritma C4.5 merupakan metode mining data yang direkomendasikan untuk menghitung daftar kelulusan pada lembaga kursus tersebut. Algoritma C4.5 yaitu hasil dari pengembangan algoritma ID3 (Iterative Dichotomiser) yang dikembangkan oleh J. Ross Quinlan pada tahun 1979 [3]. Algoritma C4.5 ini digunakan dalam sebuah klarifikasi yang menggunakan pohon keputusan [20]. Algoritma ini memiliki kelebihan yaitu pada hasil yang bermodel pohon, memiliki tingkat akurasi yang dapat diterima sehingga memudahkan dalam pemahaman algoritmanya.

Algoritma C4.5 mempunyai prinsip dasar hampir sama dengan algoritma ID3, tetapi algoritma ini memiliki beberapa perbedaan yaitu [30] :

1. Dapat menangani atribut dengan tipe kontinu dan tipe diskrit [8]. Pemilihan atribut dalam algoritma pada induksi *decision 3* adalah dengan menggunakan ukuran besar berdasarkan *entropy* yang sering dikenal dengan sebutan *information gain* sebagai bentuk *heuristic* agar dapat memilih atribut yang juga merupakan bagian dalam terbaik dari contoh ke dalam sebuah data atau kelas [9]. Semua atribut tersebut mempunyai sifat kategori yang mempunyai nilai diskrit. Atribut yang mempunyai nilai *continuous* harus dilakukan sebuah didiskritkan. Diskritisasi sebuah atribut memiliki tujuan untuk mempermudah pengelompokan data atau nilai yang dibuat berdasarkan kriteria yang telah disiapkan atau ditetapkan. Hal ini memiliki tujuan untuk menyederhanakan masalah atau objek penelitian dan meningkatkan persentase akurasi dalam proses pembelajaran
2. Dapat menyelesaikan masalah atribut yang kosong (*missing value*), nilai kosong yang ada pada sebuah *dataset* harus diisi terlebih dahulu sebelum melakukan proses pada tahap *machine learning* atau ke dalam bentuk model pada sebuah *decision tree*. Cara yang paling mudah dalam pengisian data atau atribut yang kosong adalah dengan mengisi dan memberikan nilai-nilai berdasarkan nilai yang paling sering muncul atau paling banyak dan dominan dalam atribut yang ada pada data tersebut.
3. Pemangkasan pada pohon keputusan. Saat melakukan pembangunan pohon keputusan, semakin banyak cabang yang dimiliki pohon keputusan maka hal tersebut akan mencerminkan apakah di situ ada sebuah *noise* atau *outlier* pada sebuah *training data*. Pemangkasan yang sedang dilakukan pada pohon tersebut perlu dilakukan untuk mengenali dan menghilangkan atau menghapus cabang-cabang yang ada pada data tersebut. Semakin kecil pohon yang dipangkas, dan dibuat lebih kecil lagi maka akan membuat data menjadi lebih mudah untuk dapat dipahami. Pohon semacam itu biasanya dapat menjadi lebih cepat dan lebih baik untuk melakukan dalam hal klasifikasi

4. Pemilihan pada atribut yang dilakukan dengan menggunakan *gain ratio*. *Information gain* pada ID3 biasanya mengutamakan pengujian yang akan menghasilkan banyak sekali keluaran. Dengan begitu, pada sebuah atribut yang memiliki banyak nilai-nilai yang akan dipilih sebagai *splitting* atribut.

Secara umum terdapat alur proses pada metode algoritma C4.5 saat membuat pohon keputusan dalam data mining yaitu sebagai berikut (Mahmud, et al, 2014):

1. Memilih atribut kemudian jadikan sebagai akarnya.
2. Buatlah tiap-tiap nilai di setiap cabang.
3. Bagilah kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama<sup>[7]</sup>.

Rumus algoritma C4.5 yang digunakan pada pencarian nilai Entropi pada atribut sebagai berikut :

$$\text{Entropi } (S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dan untuk pencarian nilai Gain pada atribut menggunakan rumus sebagai berikut :

$$\text{Gain } (S, A) = \text{entropi } (S) - \sum_i^n = \frac{|S_i|}{|S|} * \text{entropi } (S_i)$$

Keterangan :

- |                |                                     |
|----------------|-------------------------------------|
| S              | : Himpunan kasus                    |
| A              | : Atribut                           |
| n              | : Jumlah partisi atribut S          |
| p <sub>+</sub> | : Proporsi bernilai mendukung       |
| p <sub>-</sub> | : Proporsi bernilai tidak mendukung |
| p <sub>i</sub> | : Proporsi ke-i                     |
| S <sub>i</sub> | : Jumlah kasus ke-i                 |
| S              | : Jumlah kasus ke S                 |

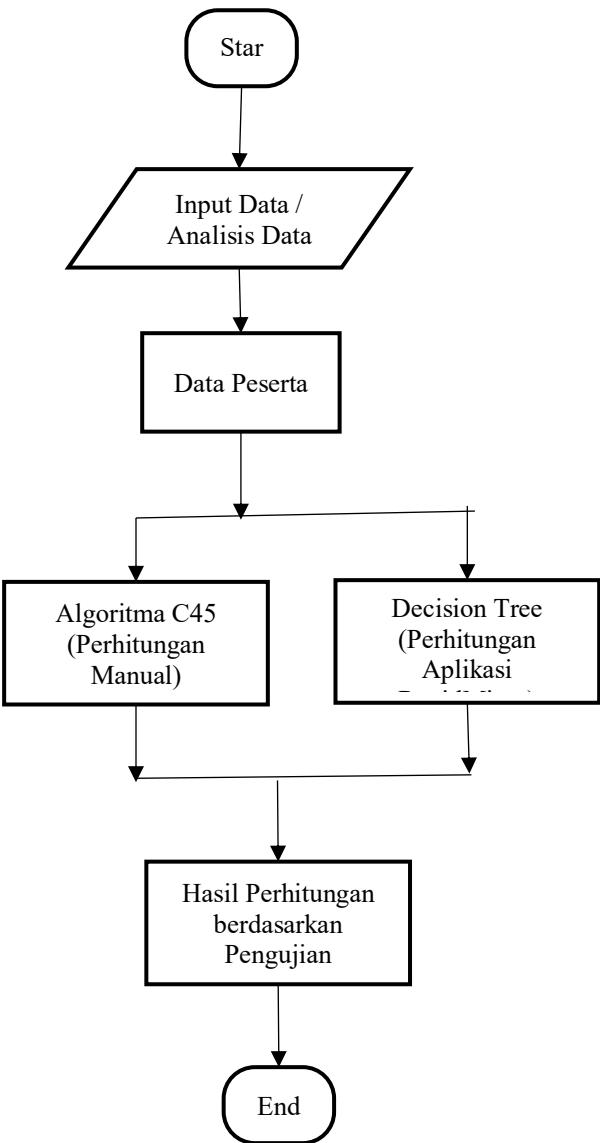
### C. Perancangan

Perancangan Algoritma C4.5 menggunakan 2 metode perhitungan, yang pertama perhitungan secara manual dan perhitungan di sebuah aplikasi yaitu *RapidMiner*. *Rapid Miner* juga sangat efektif dalam perhitungan di berbagai metode salah satunya *Decision Tree* (C4.5).

Tahapan perancangan metode yang ada di penelitian ini sebagai berikut<sup>[11]</sup>:

1. Mulai
2. Input data atau analisis data yaitu data yang mana ID, dan atributnya
3. Data peserta semua data yang berdasarkan keterkaitan penelitian
4. Memulai perhitungan, yang pertama untuk perhitungan manual digunakan rumus yang sudah dijelaskan sebelumnya. Sedangkan untuk perhitungan pada aplikasi *RapidMiner* akan secara otomatis pada aplikasi tersebut.
5. Setelah semua perhitungan selesai akan ada hasil kedua perhitungan tersebut.
6. Selesai

Proses algoritma penelitian yang diterapkan dapat dijabarkan dalam bentuk flowchart yaitu sebagai berikut :



Gambar 1 Desain Flowchart Penelitian

Pembuatan perancangan ini untuk memudahkan dalam melakukan pembuatan jurnal agar susunan urutannya menyesuaikan dengan tepat.

#### D. Pengolahan Data

Data yang akan digunakan yaitu tentang pendaftaran pada kursus komputer dan mobil. Data pembahasan ini menggunakan tabel Atribut yaitu [1]:

- No.induk (sebagai ID),
  - Tempat tanggal Lahir (Luar Sampit, Sampit, Kotim),
  - Alamat (Hm.Arsyad, Baamang, Sudirman), dan
  - Umur (Lebih dari 15 tahun, Kurang dari 15 tahun).
- Untuk rincian informasi data lebih jelas di bawah ini :

TABEL I  
INFORMASI DATA

No	Atribut	Nilai Atribut	Keterangan
1	No. Induk		ID
2	Tempat Tanggal Lahir	Sampit	Kota yang berada di sekitar sampit
		Kotawaringin Timur	Semua Kota dan daerah yang mencakup Kotim
		Luar Sampit	Kota yang berada di luar Sampit
3	Alamat	Hm.Arsyad	Area rumah disekitar jalan Hm.Arsyad
		Baamang	Area rumah disekitar jalan Baamang
		Sudirman	Area rumah di sekitar jalan Sudirman
4	Umur	< 30 Tahun	yang mendaftar di bawah 30 tahun
		> 30 Tahun	Yang mendaftar di umur lebih dari 30 tahun

### III. HASIL DAN PEMBAHASAN

#### A. Pengolahan Data

Data dikumpulkan kemudian dikelola, pada tahap ini dilakukan perhitungan atribut-atribut seperti nama siswa atau nama pendaftar [1],[3]. Tahap sebelumnya telah dilakukan perancangan data yang diperoleh sehingga dapat dilanjutkan ke tahap berikutnya. Tahap selanjutnya perhitungan pada entropi beserta gain untuk melakukannya dilihat dari jumlah data (jumlah keseluruhan data, dari tabel 2 dan juga tabel 3), pada peserta lulus tepat waktu dan peserta tidak lulus tepat waktu [29]. Setelah mendapatkan hasil dari Gain, langsung ke pohon keputusan pada setiap tabel (mengemudi dan komputer). Tabel di atas juga sudah dijelaskan analisis perancangan data nya, yang pertama perhitungan pada tabel 2 (Mengemudi) sebagai berikut :

TABEL II  
PERHITUNGAN MENGENUDI

	Atribut	Jumlah Data	Lulus Tepat Waktu	Tidak Lulus Tepat Waktu	Entropi	Gain
Node						
Total		317	302	15	0.1348626	
Tempat Tanggal Lahir						
Sampit		30	20	10	0.9182958	0.0288982
Kotawaringin Timur		280	280	0	0	
Luar Sampit		7	2	5	0.8631206	
Alamat						
Hm.Arsyad		277	277	0	0	0.1025091
Baamang		7	5	2	0.8631206	
Sudirman		33	20	13	0.1277038	
Umur						
<30		287	287	0	0	0.0402254
>30		30	15	15	1	

Rumus yang digunakan pada pencarian nilai Entropi pada atribut Tempat Tanggal Lahir sebagai berikut :

$$\text{Entropi (S)} = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dan untuk pencarian nilai Gain pada atribut Tempat Tanggal Lahir menggunakan rumus sebagai berikut :

$$\text{Gain (S,A)} = \text{entropi (S)} - \sum_i^n \frac{|S_i|}{|S|} * \text{entropi (S)}_i$$

Keterangan :

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut S
- p+ : Proporsi bernilai mendukung
- p- : Proporsi bernilai tidak mendukung
- p<sub>i</sub> : Proporsi ke-i
- |S<sub>i</sub>| : Jumlah kasus ke-i
- |S| : Jumlah kasus ke S

Yang pertama dihitung yaitu Entropi Total Mengemudi :

Entropi (Total)

=

$$\sum_{i=1}^n \left( -\frac{302}{317} * \log_2 \left( \frac{302}{317} \right) \right) + \left( -\frac{15}{317} * \log_2 \left( \frac{15}{317} \right) \right) = 0.1348626$$

Yang kedua menghitung Entropi atribut Tempat Tanggal Lahir :

Entropi (Tempat Tanggal Lahir Daerah Sampit)

$$= \sum_{i=1}^n \left( -\frac{20}{30} * \log_2 \left( \frac{20}{30} \right) \right) + \left( -\frac{10}{30} * \log_2 \left( \frac{10}{30} \right) \right) = 0.9182958$$

Entropi (Tempat Tanggal Lahir Daerah Kotawaringin Timur)

$$= \sum_{i=1}^n \left( -\frac{280}{280} * \log_2 \left( \frac{280}{280} \right) \right) + \left( -\frac{0}{280} * \log_2 \left( \frac{0}{280} \right) \right) = 0$$

Entropi (Tempat Tanggal Lahir Daerah Luar Sampit)

$$= \sum_{i=1}^n \left( -\frac{2}{7} * \log_2 \left( \frac{2}{7} \right) \right) + \left( -\frac{5}{7} * \log_2 \left( \frac{5}{7} \right) \right) = 0.8631206$$

Menghitung Gain mengemudi pada Tempat Tanggal Lahir  
Gain = **0.1348626** - ((30/317 \* 0.9182958) + (280/317 \* 0) + (7/317 \* 0.8631206))  
= 0.0288982

Yang ketiga Entropi atribut Alamat :

Entropi (Alamat Jalan Hm.Arasyad)

$$= \sum_{i=1}^n \left( -\frac{277}{277} * \log_2 \left( \frac{277}{277} \right) \right) + \left( -\frac{0}{277} * \log_2 \left( \frac{0}{277} \right) \right) = 0$$

Entropi (Alamat Jalan Baamang)

$$= \sum_{i=1}^n \left( -\frac{5}{7} * \log_2 \left( \frac{5}{7} \right) \right) + \left( -\frac{2}{7} * \log_2 \left( \frac{2}{7} \right) \right) = 0.8631206$$

Entropi (Alamat Jalan Sudirman)

$$= \sum_{i=1}^n \left( -\frac{20}{33} * \log_2 \left( \frac{20}{33} \right) \right) + \left( -\frac{13}{33} * \log_2 \left( \frac{13}{33} \right) \right) = 0.1277038$$

Menghitung Gain mengemudi pada Alamat

Gain = **0.1348626** - ((277/317 \* 0) + (7/317 \* 0.8631206) + (33/317 \* 0.1277038))  
= 0.1025091

Yang keempat menghitung Entropi atribut Umur

Entropi (Umur di bawah 30 Tahun)

$$= \sum_{i=1}^n \left( -\frac{287}{287} * \log_2 \left( \frac{287}{287} \right) \right) + \left( -\frac{0}{287} * \log_2 \left( \frac{0}{287} \right) \right) = 0$$

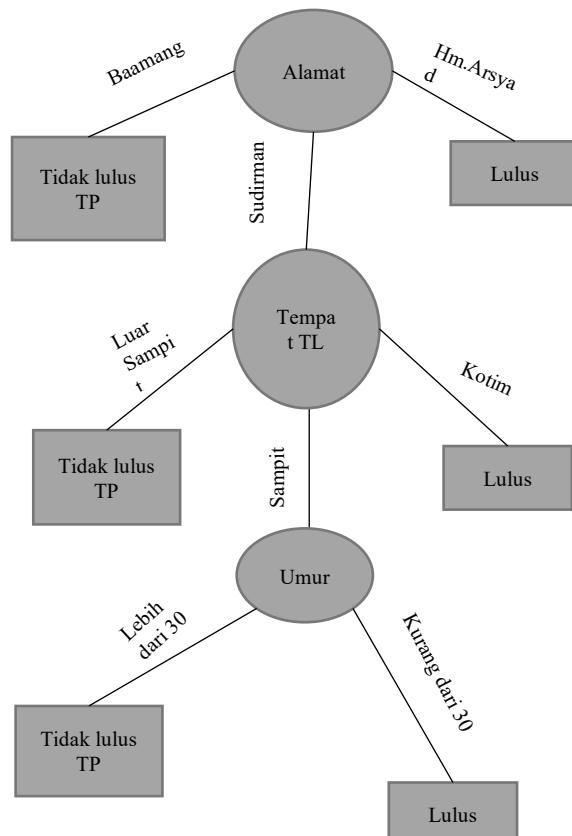
Entropi (Umur di atas 30 tahun)

$$= \sum_{i=1}^n \left( -\frac{15}{30} * \log_2 \left( \frac{15}{30} \right) \right) + \left( -\frac{15}{30} * \log_2 \left( \frac{15}{30} \right) \right) = 1$$

Menghitung Gain mengemudi pada Umur

$$\text{Gain} = \text{0.1348626} - ((287/317 * 0) + (30/317 * 1)) = 0.0402254$$

Dari hasil perhitungan nilai entropi, gain pada tabel 2 (mengemudi) maka dapat disimpulkan nilai hasil gain tertinggi di pada tabel 2 (mengemudi) yaitu gain alamat yang mendapatkan hasil 0.1025091 untuk menentukan nilai akarnya, dan setelah diperoleh maka hasil dari pohon keputusan sebagai berikut :



Gambar 2 Pohon Keputusan

Kesimpulan Pohon keputusan menunjukkan bahwa yang paling cepat lulus tepat waktu yaitu peserta dengan alamat Hm.Arsyad, tempat tanggal lahir di Kotawaringin Timur dan berumur sekitar kurang dari 30 tahun. Sebaliknya peserta yang tidak lulus tepat waktu di sekitar alamat Baamang, tempat tanggal lahir di Luar Sampit, dan berumur lebih dari 30 tahun. Berlaku pada peserta kursus mengemudi. Tahap selanjutnya perhitungan tabel 3 Komputer sebagai berikut :

TABEL III  
PERHITUNGAN KOMPUTER

Atribut	Jumlah data	Lulus Tepat Waktu	Tidak Lulus Tepat Waktu	Entropi	Gain
Node					
Total	77	61	16	0.7372414	
<b>Tempat Tanggal Lahir</b>					
Sampit	21	20	1	0.2761954	0.0521896
Kotawaringin Timur	0	0	0	0	
Luar Sampit	56	41	15	0.8383729	
<b>Alamat</b>					
Hm.Arsyad	43	39	4	0.4464813	0.203890
Baamang	22	10	12	0.9940302	
Sudirman	12	12	0	0	
<b>Umur</b>					
<30	46	33	13	0.8589810	0.0394182
>30	31	28	3	0.4586858	

Rumus yang digunakan pada pencarian nilai Entropi pada atribut Tempat Tanggal Lahir sebagai berikut :

$$\text{Entropi } (S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Dan untuk pencarian nilai Gain pada atribut Tempat Tanggal Lahir menggunakan rumus sebagai berikut :

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

Keterangan :

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut S
- p<sub>+</sub> : Proporsi bernilai mendukung
- p<sub>-</sub> : Proporsi bernilai tidak mendukung
- p<sub>i</sub> : Proporsi ke-i
- |S<sub>i</sub>| : Jumlah kasus ke-i
- |S| : Jumlah kasus ke S

Yang pertama dihitung yaitu Entropi Total Komputer :

$$\begin{aligned} &\text{Entropi} && (\text{Total}) \\ &= \sum_{i=1}^n \left( -\frac{61}{77} * \log_2 \left( \frac{61}{77} \right) \right) + \left( -\frac{16}{77} * \log_2 \left( \frac{16}{77} \right) \right) = 0.7372414 \end{aligned}$$

Yang kedua menghitung Entropi atribut Tempat Tanggal Lahir :

$$\text{Entropi } (\text{Tempat Tanggal Lahir Daerah Sampit}) =$$

$$\begin{aligned} &= \sum_{i=1}^n \left( -\frac{20}{21} * \log_2 \left( \frac{20}{21} \right) \right) + \left( -\frac{1}{21} * \log_2 \left( \frac{1}{21} \right) \right) = 0.2761954 \end{aligned}$$

Entropi (Tempat Tanggal Lahir Daerah Kotawaringin Timur)

$$= \sum_{i=1}^n \left( -\frac{0}{0} * \log_2 \left( \frac{0}{0} \right) \right) + \left( -\frac{0}{0} * \log_2 \left( \frac{0}{0} \right) \right) = 0$$

Entropi (Tempat Tanggal Lahir Daerah Luar Sampit)

$$\begin{aligned} &= \sum_{i=1}^n \left( -\frac{41}{56} * \log_2 \left( \frac{41}{56} \right) \right) + \left( -\frac{15}{56} * \log_2 \left( \frac{15}{56} \right) \right) = 0.8383729 \end{aligned}$$

Menghitung Gain Komputer pada Tempat Tanggal Lahir  
 $\text{Gain} = 0.7372414 - ((21/77 * 0.2761954) + (0/77 * 0) + (56/77 * 0.8383729))$   
 $= 0.0521896$

Yang ketiga Entropi atribut Alamat :

$$\text{Entropi } (\text{Alamat Jalan Hm.Arsyad}) =$$

$$\begin{aligned} &= \sum_{i=1}^n \left( -\frac{39}{43} * \log_2 \left( \frac{39}{43} \right) \right) + \left( -\frac{4}{43} * \log_2 \left( \frac{4}{43} \right) \right) = 0.4464813 \end{aligned}$$

Entropi (Alamat Jalan Baamang)

$$\begin{aligned} &= \sum_{i=1}^n \left( -\frac{10}{22} * \log_2 \left( \frac{10}{22} \right) \right) + \left( -\frac{12}{22} * \log_2 \left( \frac{12}{22} \right) \right) = 0.9940302 \end{aligned}$$

Entropi (Alamat Jalan Sudirman)

$$\begin{aligned} &= \sum_{i=1}^n \left( -\frac{12}{12} * \log_2 \left( \frac{12}{12} \right) \right) + \left( -\frac{0}{12} * \log_2 \left( \frac{0}{12} \right) \right) = 0 \end{aligned}$$

Menghitung Gain Komputer pada Alamat

$$\begin{aligned} \text{Gain} &= 0.7372414 - ((43/77 * 0.4464813) + (22/77 * 0.9940302) + (12/77 * 0)) \\ &= 0.2038990 \end{aligned}$$

Yang keempat menghitung Entropi atribut Umur :

$$\text{Entropi } (\text{Umur dibawah 30 Tahun})$$

$$= \sum_{i=1}^n \left( -\frac{33}{46} * \log_2 \left( \frac{33}{46} \right) \right) + \left( -\frac{13}{46} * \log_2 \left( \frac{13}{46} \right) \right) = 0.8589810$$

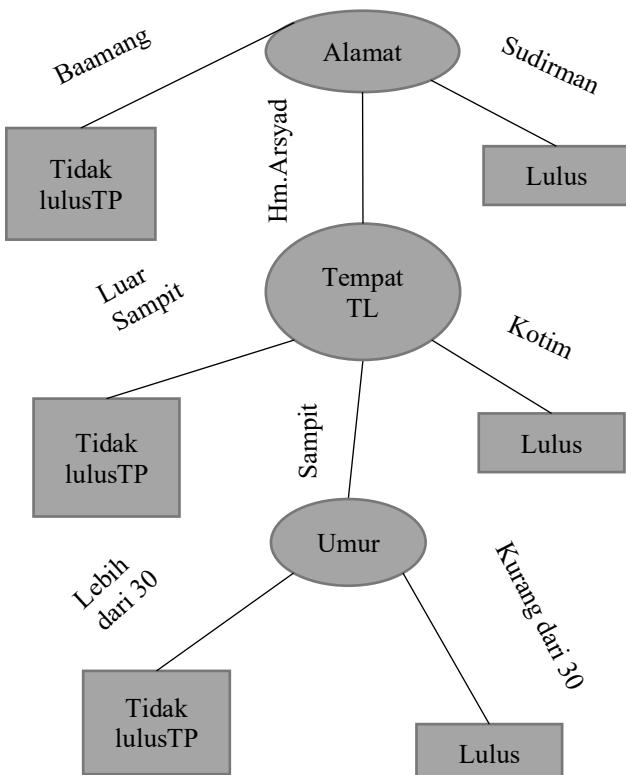
Entropi (Umur di atas 30 tahun)

$$= \sum_{i=1}^n \left( -\frac{28}{31} * \log_2 \left( \frac{28}{31} \right) \right) + \left( -\frac{3}{31} * \log_2 \left( \frac{3}{31} \right) \right) = 0.4586858$$

Menghitung Gain komputer pada Umur

$$\text{Gain} = 0.7372414 - ((46/77 * 0.8589810) + (31/77 * 0.4586858)) \\ = 0.0394182$$

Dari hasil perhitungan nilai entropi, gain pada tabel 3 (komputer) maka dapat disimpulkan nilai hasil gain tertinggi di pada tabel 3 (komputer) yaitu gain alamat yang mendapatkan hasil 0.2038990 untuk menentukan nilai akarnya, dan setelah diperoleh maka hasil dari pohon keputusan sebagai berikut :



Gambar 3 Pohon Keputusan

Kesimpulan Pohon keputusan menunjukkan bahwa yang paling cepat lulus tepat waktu yaitu peserta dengan alamat Sudirman, tempat tanggal lahir di Kotawaringin Timur dan berumur sekitar kurang dari 30 tahun. Sebaliknya peserta yang tidak lulus tepat waktu di sekitar alamat Baamang, tempat tanggal lahir di Luar Sampit, dan

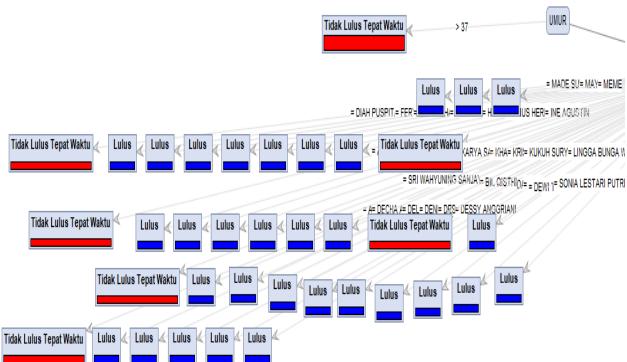
berumur lebih dari 30 tahun. Berlaku pada peserta kursus komputer.

#### B. Rapid Miner

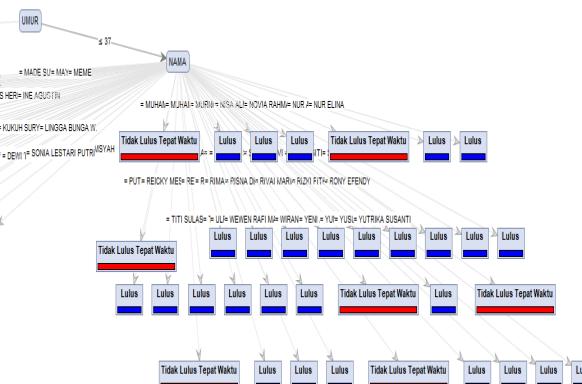
Rapid Miner merupakan aplikasi perangkat lunak yang dikembangkan oleh perusahaan [23], [27]. Pada aplikasi rapid miner terdapat sejarah yaitu sebelumnya telah dikenal YALE (Yet Another Learning Environment), yang dikembangkan oleh Ralf Klinkenberg, Indo Mierswa, dan Simon Fischer pada tahun 2001 dari unit Kecerdasan Buatan Universitas Teknik Dortmund [6], [12]. Rapid Miner juga sangat efektif dalam perhitungan di berbagai metode salah satunya *Decision Tree* (C4.5).

#### 1. Pengujian Metode Decision Tree (C4.5)

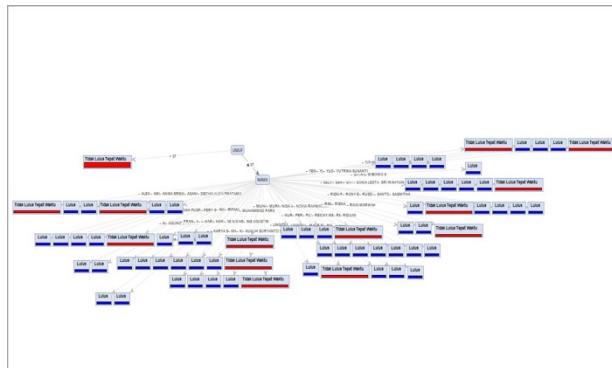
Pengujian Metode Decision Tree (C4.5) pada tabel 3 komputer. Data yang diambil merupakan total keseluruhan data pada tabel komputer sebanyak 77 peserta yang dibagi menjadi 61 untuk peserta Lulus tepat waktu dan 16 untuk peserta Tidak lulus tepat waktu. Untuk melakukan perhitungan seperti nilai Entropi dan Gain dengan cara yang sama pada aplikasi Rapid Miner, akan memperoleh pohon keputusan sebagai berikut :



Gambar 4 Pohon Keputusan

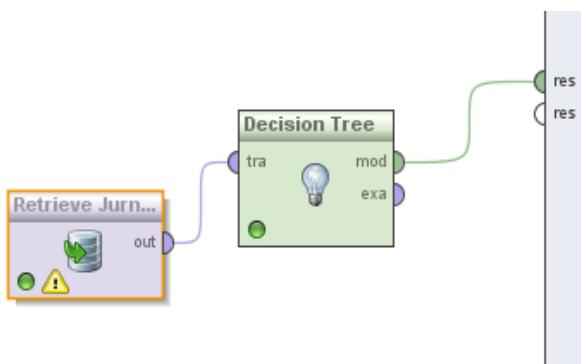


Gambar 5 Pohon Keputusan



Gambar 6 Pohon Keputusan

Sebelum mendapat pohon keputusan pada aplikasi Rapid Miner, data yang diperoleh di analisis ke sebuah excel yang kemudian di import setelah itu sambungkan output yang sudah jadi ke decision tree pada aplikasi Rapid Miner. Tampilannya akan memperoleh hasil pohon keputusan setelah di jalankan, bentuk gambaran di bawah ini merupakan hasil sebelum di jalan menjadi pohon keputusan.



Gambar 7 Validation Model Decision Tree

Pengujian terhadap aplikasi *Rapid Miner* ini untuk mendapatkan hasil pohon keputusan dengan menggunakan

gain ratio pada *decision tree* dengan hasil yang sama akuratnya dengan perhitungan manual.

#### IV. EVALUASI

Dari hasil pohon keputusan kedua tabel mengemudi dan komputer terdapat jumlah hasil kelulusan minat masyarakat pada bidang kursus mengemudi dan komputer. Beserta hasil pohon keputusan pada aplikasi RapidMiner juga terdapat hasil kelulusan yang akurat. Berdasarkan uji algoritma C45 dan aplikasi RapidMiner memiliki tingkat akurasi kurang dari 100%. Akurasi dilihat dari Lulus dan Tidak Lulus. Selain itu juga perhitungan pada aplikasi Rapid Miner yang telah dilakukan pada tabel komputer, kesimpulan akhir data banyak peserta yang lulus tepat waktu sekitar 70% sedangkan yang tidak lulus tepat waktu hanya 30%.

#### V. KESIMPULAN

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa :

Penerapan Algoritma C45 oleh pohon keputusan untuk memprediksi jumlah seberapa banyak minat masyarakat berdasarkan data kelulusan dan perhitungan ketepatan lulus peserta pendaftar.

Hasil dari penerapan Algoritma C45 oleh pohon keputusan untuk memprediksi jumlah kelulusan pertahun pada kursus mengemudi dan komputer berdasarkan Tabel pohon keputusan dapat juga membantu dalam menentukan pendaftar berikutnya sesuai dengan kebutuhan dan keinginan.

Sedangkan hasil penerapan oleh Aplikasi RapidMiner juga didapatkan banyak pemminat masyarakat untuk kursus mengemudi ataupun komputer setiap tahun nya. Dilihat dari jumlah kelulusan lebih banyak daripada jumlah yang tidak tepat waktu untuk lulus.

#### TERIMA KASIH

Dalam pembuatan dan penggeraan jurnal ini penulis menyadari bahwa selesai penggeraan tidak lepas dari dukungan beserta bimbingan dari dosen maupun pihak lainnya. Oleh karena itu penulis menyampaikan ucapan terima kasih kepada pihak :

1. Bapak Nurrahman, S.Kom. M.M selaku dosen mata kuliah penambangan data yang memberikan arahan pembuatan jurnal ini.
2. Bapak Lukman Bachtiar, S.Kom., MM selaku pembimbing yang telah memberikan masukan dalam menyusun jurnal ini.
3. Rekan dari seangkatan Sistem Informasi Fakultas Ilmu Komputer Universitas Darwan Ali Sampit

## DAFTAR PUSTAKA

- [1] Astuti, Indah Puji. 2017. Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5. *Fountain of Informatics Journal* Vol. 2 No. 2. Ponorogo: Universitas Muhammadiyah Ponorogo
- [2] Ahmad, Hermansyah Nur, Vincent Suhartono, dan Ika Novita Dewi. 2017. Penentuan Tingkat Kelulusan Tepat Waktu Mahasiswa SMTIK Subang Menggunakan Algoritma C4. 5. *Jurnal Teknologi Informasi* Vol. 13 No. 1. Universitas Dian Nuswantoro
- [3] Azwanti, Nurul. 2018. Algoritma C4. 5 Untuk Memprediksi Mahasiswa yang Mengulang Mata Kuliah (Studi Kasus di Amik Labuhan Batu). *Jurnal SIMETRIS* Volume 9 Nomor 1. Batam: Universitas Putera Batam
- [4] Azwanti, Nurul. 2018. Algoritma C4. 5 Untuk Memprediksi Penjualan Motor pada PT. Capella Dinamik Nusantara Cabang Muka Kuning. *Jurnal Ilmian Ilmu Komputer* Volume 13 Nomor 1. Batam: Universitas Putera Batam
- [5] Hartina, Ayu, Takdir Tahir, Nurhaya Nurdin, dan Midawati Djafar. 2017. Faktor yang Berhubungan dengan Kelulusan Uji Kompetensi Ners Indonesia (UKNI) di Regional Sulawesi. *JPPNI* Volume 2 Nomor 2. Makassar: Universitas Hasanuddin
- [6] Jura, Suwatri, Sitti Zuhriyah, dan A Edeth Fuari. 2017. Sistem Pendukung Keputusan Penentuan Kelulusan Hasil Ujian Akhir Sekolah Menggunakan Metode Simple Additive Weighting. *JTRISTE* Volume 4 Nomor 1. Makassar: STMIK Handayani Makassar
- [7] Kristanto, Nanang. 2018. Desain Kurikulum Kursus dan Pelatihan Komputer (Aplikasi Perkantoran) Pada Sanggar Kegiatan Belajar (SKB) untuk Meningkatkan mutu lulusan di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Pendidikan Luar Sekolah* Volume 2 Nomor 2. Balai Pengembangan Pendidikan Anak Usia Dini Dan Pendidikan Masyarakat Papua
- [8] Kartini, Dwi dan Muliadi. 2016. Rancang Bangun Aplikasi K-Means untuk Klasifikasi Kelulusan Siswa Sekolah Kependidikan Negara Daerah Kalimantan Selatan. *Jurnal ProTekInfo* Volume 3 Nomor 1. Banjarmasin: Universitas Lambung Mangkurat
- [9] Kurniawan, Angga. 2019. Implementasi Data Mining Algoritma C4.5 Untuk Memprediksi Kelulusan Uji Kompetensi SMK Teknik Komputer dan Jaringan (TKJ) (Study Kasus: SMK Pembangunan Daerah Lubuk Pakam). *Jurnal Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)* Volume 7 Nomor 1. Medan: STMIK Budi Dharma
- [10] Mustaqbal, M. Sidi, Roeri Fajri Firdaus, dan Hendra Rahmadi. 2015. Pengujian Aplikasi Menggunakan Black Box Testing Boundary Value Analysis (Studi Kasus: Aplikasi Prediksi Kelulusan SNMPTN). *Jurnal Ilmiah Teknologi Informasi Terapan* Voolume 1 Nomor 3. Bandung: Universitas Widyatama
- [11] Meilani, Budanis Dwi dan Nofi Susanti. 2015. Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa dengan Metode Naive Bayes. *Jurnal Ilmiah NERO* Volume 1 Nomor 3. Surabaya: Institut Teknologi Adhi Tama Surabaya
- [12] Mardalius. 2018. Pemanfaatan Rapidminer Studio 8.2 untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *Jurnal Teknologi dan Sistem Informasi* Volume IV Nomor 2. Sumatera Utara: STMIK Royal Kisaran
- [13] Muna, Fauzul. 2017. Implementasi Metode Naive Bayes untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama. Skripsi. Semarang: Universitas Diponegoro
- [14] Novita, Rina. 2016. Teknik Data Mining: Algoritma C 4.5. Komunitas eLearning IlmuKomputer.com
- [15] Putri, Rana Puspita Sari dan Indra Waspada. 2018. Penerapan Algoritma C4. 5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. *Jurnal Ilmu Komputer dan Informatika* Vol. 4 No. 1. Semarang: Universitas Diponegoro
- [16] Pratama, Arif, Randy Cahya Wihandika, dan Dian Eka Ratnawati. 2018. Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* Volume 2 Nomor 4. Malang: Universitas Brawijaya
- [17] Ridwan, Mujib. 2017. Sistem Rekomendasi Proses Kelulusan Mahasiswa Berbasis Algoritma C4. 5. *Jurnal Ilmiah Informatika* Vol. 2 No. 1. Surabaya: Universitas Islam Negeri Sunan Ampel
- [18] Rohman, Abdul dan Muhammad Rochman. 2019. Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. *Jurnal Neo Teknika* Volume 5 Nomor 1.
- [19] Ridwan, Mujib. 2017. Sistem Rekomendasi Proses Kelulusan Mahasiswa Berbasis Algoritma Klasifikasi C4. 5. *Jurnal Ilmiah Informatika* Volume 2 Nomor 1. Surabaya: UIN Sunan Ampel Surabaya
- [20] Rismayanti. 2016. Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan

Medan. Jurnal Media Infotama Volume 12 Nomor 2. Medan: Sekolah Tinggi Teknik Harapan Medan Jurnal Positif Volume 3 Nomor 1. Malang: STMIK Asia Malang

[21] Rohmawan, Eko Prasetyo. 2018. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network. Jurnal Ilmiah MATRIK Volume 20 Nomor 1. Bengkulu: Universitas Dehasen Bengkulu

[22] Sari, Riyani Wulan, Anjar Wanto, dan Perdama Windarto. 2018. Implementasi Rapidminer dengan Metode K-Means (Study Kasus: Imunisasi Campak pada Balita Berdasarkan Provinsi). KOMIK Volume 2 Nomor 1. Pematangsiantar: STIKOM Tunas Bangsa

[23] Silalahi, Mesri. 2018. Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk pada PT. Batamas Niaga Jaya. Computer Based Information System Journal Volume 6 Nomor 2. Batam: Universitas Putera Batam

[24] Salmu, Supardi dan Ahmad Solichin. 2017. Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. Prosiding Seminar Nasional Multidisiplin Ilmu. Jakarta: Universitas Budi Luhur

[25] Suhardjono, Gandha Wijaya, dan Abdul Hamid. 2019. Prediksi Waktu Kelulusan Mahasiswa Menggunakan SVM Berbasis PSO. Biangalal Informatika Volume 7 Nomor 2. Universitas Bina Sarana Informatika

[26] Sinaga, Firman Mangasi F. 2017. Impementasi Jaringan Syaraf Tiruan Metode Backpropagation untuk prediksi Kelulusan Siswa Sekolah Menengah Atas (Studi Kasus: SMA Cahaya Medan). Majalah Ilmiah INTI Volume XII Nomor 1. Medan: STMIK Budidarma Medan

[27] Wikipedia.com. (2020, 17 April). RapidMiner. Diakses pada 15 Agustus 2020, dari <https://id.m.wikipedia.org/wiki/RapidMiner>

[28] Wahyuni, Sri, Kana Saputra S., dan Mochammad Iswan Perangin-angin. 2017. Implementasi Rapidminer dalam Menganalisa Data Mahasiswa Drop Out. Medan: Universitas Pembangunan Pancabudi Meda

[29] Yalidhan, Muhammad Dedeck dan Muhammad Faisal Amin. 2018. Imolementasi Algoritma Backpropagation untuk Memprediksi Kelulusan Mahasiswa. Kumpulan Jurnal Ilmu Komputer (KJIK) Volume 5 Nomor 2. Banjarbaru: STMIK Banjarbaru

[30] Yosepta, Dwi Anggih dan Tria Aprilianto. 2017. Analisa Pola Kelulusan Mahasiswa pada Sekolah Tinggi Manajemen Informatika & Komputer Asia Malang dengan Menggunakan Algoritma Interactive Dichotomiser 3 (ID3).