

Analisa Faktor Kesulitan Mahasiswa Dalam Menyelesaikan Skripsi Dengan Algoritma C4.5

Azi arisandi¹, Irfan Sudahri Damanik², Ilham Syahputra Saragih³
^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar – Indonesia

Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara

¹aziarisandi111@gmail.com, ²Irfansudahri@stikomtunasbangsa.com,
³ilhamsyahputra@amiktunasbangsa.ac.id

Abstract

The purpose of this study was to determine the most dominant factor in students' difficulties in completing their thesis. In this study, researchers used data mining techniques using the C4.5 algorithm. Sources of research data used were obtained from interviews and questionnaires to students randomly in the STIKOM Tunas Bangsa Pematangsiantar environment. The research variables used were motivation to graduate on time (C1), ability to write scientific papers (C2), availability of learning resources (C3), guidance process (C4), and peer environment (C5). In this study, the alternative used as a sample is the student of STIKOM Tunas Bangsa, semester eight. The results of the classification using the C4.5 algorithm and testing with rapidminer software found that the most dominant factor in student difficulty in completing theses was the motivation to graduate on time (C1) with a gain value of 0.204593531.

Keywords : C4.5 Algorithm, Thesis, Data Mining, classification

Abstrak

Tujuan dari penelitian ini adalah untuk mengetahui faktor yang paling dominan terhadap kesulitan mahasiswa dalam menyelesaikan skripsi. Pada penelitian ini peneliti menggunakan teknik Data Mining dengan menggunakan algoritma C4.5. Sumber data penelitian yang digunakan diperoleh dari wawancara dan pemberian angket kepada mahasiswa secara acak di lingkungan STIKOM Tunas Bangsa Pematangsiantar. Adapun variabel penelitian yang digunakan adalah motivasi lulus tepat waktu (C1), kemampuan menulis karya ilmiah (C2), ketersediaan sumber belajar (C3), proses bimbingan (C4), dan lingkungan teman sebaya (C5). Pada penelitian ini alternatif yang digunakan sebagai sampel adalah Mahasiswa STIKOM Tunas Bangsa, semester delapan. Hasil dari klasifikasi menggunakan algoritma C4.5 dan pengujian dengan software rapidminer didapatkan faktor yang paling dominan kesulitan mahasiswa menyelesaikan skripsi adalah motivasi lulus tepat waktu (C1) dengan nilai gain sebesar 0,204593531.

Kata Kunci : Algoritma C4.5, Skripsi, Data Mining, klasifikasi

1. Pendahuluan

Pendidikan, penelitian dan pengabdian kepada masyarakat merupakan Tri Dharma yang dijunjung oleh seluruh perguruan tinggi di Indonesia. Perguruan tinggi dilaksanakan tidak lain sebagai lembaga yang menangani bidang penelitian. Dalam rangka memenuhi tuntutan penyelenggaraan penelitian, maka para mahasiswa didorong untuk melakukan penelitian sebagai syarat utama dalam menyelesaikan studinya, yaitu menyusun skripsi. Menurut [1] skripsi adalah sebuah karya ilmiah yang dibuat berdasarkan pengetahuan-pengetahuan khusus dan fakta-fakta yang jelas. Dengan menyusun skripsi mahasiswa diharapkan dapat memecahkan permasalahan yang dihadapi secara ilmiah, serta mengembangkan wawasannya menjadi lebih luas.

Besarnya jumlah mahasiswa STIKOM Tunas Bangsa yang melakukan perpanjangan SK bimbingan dianggap mengalami kesulitan dalam menyelesaikan skripsi yang disebabkan oleh beberapa faktor seperti motivasi lulus tepat waktu, kemampuan menulis karya ilmiah, ketersediaan sumber belajar, proses bimbingan, lingkungan teman sebaya. Dengan demikian, pihak kampus harus membuat suatu kebijakan untuk mengurangi jumlah mahasiswa yang melakukan perpanjangan SK bimbingan dengan mendeteksi data mahasiswa yang berpotensi mengalami kesulitan dalam menyelesaikan skripsi dan kemudian mencari faktor apa yang paling dominan dalam kesulitan mahasiswa menyelesaikan skripsi.

Dalam mengetahui faktor yang paling dominan terhadap kesulitan mahasiswa menyelesaikan skripsi dapat menggunakan hasil klasifikasi data mahasiswa. Klasifikasi merupakan kegiatan dalam mengekstrak dan memprediksi kategori untuk masing-masing data. Salah satu algoritma klasifikasi yang bisa digunakan adalah C4.5” [2]. Algoritma C4.5 merupakan struktur pohon yang terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. “Algoritma C4.5 memiliki kelebihan yaitu mudah dimengerti, fleksibel dan menarik karena divisualisasikan dalam bentuk gambar (pohon keputusan)” [2].

Penelitian yang dijadikan sebagai referensi dalam penelitian ini yaitu penelitian yang dilakukan oleh [3] yang berjudul “Analisa Algoritma C4.5 Dalam Menentukan Faktor Yang Mempengaruhi Munculnya Professional Blogger”. Tujuan penelitian ini adalah menentukan urutan dan pentingnya faktor-faktor yang mempengaruhi seseorang agar dapat menjadi blogger profesional. Atribut-atribut yang digunakan pada penelitian ini adalah pendidikan, kepentingan politik, topik, penggunaan media lokal dan pengaruh politik lokal dan sosial.

Penelitian ini menggunakan *decission tree* yang termasuk dalam algoritma klasifikasi data mining untuk mengetahui faktor yang paling berpengaruh terhadap kesulitan mahasiswa menyelesaikan skripsi. Dengan menggunakan algoritma C4.5 algoritma ini membuat kalsifikasi lebih sederhana, mudah dan juga tingkat akurasi yang tinggi.

2. Metodologi Penelitian

2.1. Data Mining

Data Mining berasal dari berbagai ilmu pengetahuan, yang meliputi *machine learning* atau *pattern recognition*, statistik atau kecerdasan buatan, dan sistem basis data (Muflikhah *et al.*, 2018). “Data Mining adalah proses penggalian dan pertambangan pengetahuan dari sejumlah data yang besar, *database* atau repository *database* lainnya. Tujuan utama dari penambangan data ini untuk menemukan pengetahuan baru yang tersembunyi dari *database* tersebut”[5].

2.2. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang membentuk pohon keputusan. Didalam algoritma C4.5, pohon keputusan yang dibentuk berdasarkan kriteria kerriteria yang telah ditentukan [7]. “Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan yaitu mudah dipahami, fleksibel dan menarik karena dapat diterapkan dalam bentuk gambar (pohon keputusan)” [2].

Menurut [8] ada 4 langkah yang digunakan dalam membangun pohon keputusan pada algoritma C4.5 adalah sebagai berikut :

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

3. Hasil dan Pembahasan

Dalam penelitian ini data yang digunakan akan diolah dari hasil kuisioner, yang diberikan kepada mahasiswa semester 8 dengan jumlah mahasiswa sebanyak 142. Penulis memberikan kuisioner kepada mahasiswa dengan jumlah 100 sampel data. Data yang digunakan terdiri dari :

- a) Motivasi Lulus tepat waktu (C1).
- b) Kemampuan Menulis Karya Ilmiah (C2).
- c) Ketersedian Sumber Belajar (C3).
- d) Proses Bimbingan (C4).
- e) Lingkungan Teman Sebaya (C5).

Setiap faktor atau variabel terdiri dari pertanyaan yang diajukan ke mahasiswa yang diberikan. Kuisioner yang telah dilakukan selanjutnya mencari rata-rata dari faktor atau variabel yang digunakan. Data yang digunakan merupakan jenis *statistik deskriptif* dengan mahasiswa STIKOM Tunas Bangsa. Kuisioner yang diberikan menggunakan *linker 4* yang terdiri dari SB (Sangat Baik), B (Baik), CB (Cukup Baik) dan KB (Kurang Baik).

Tabel 1. Data Penelitian yang Sudah Diolah

No	Responden	C1	C2	C3	C4	C5	Kesulitan
1	M1	SB	B	B	B	B	Tidak Sulit
2	M2	SB	C	B	B	B	Sulit
3	M3	SB	C	B	SB	B	Sulit
4	M4	SB	C	B	B	SB	Sulit
5	M5	B	B	B	B	SB	Tidak Sulit
6	M6	B	C	B	B	B	Tidak Sulit
7	M7	SB	B	B	SB	B	Tidak Sulit
8	M8	SB	C	B	B	B	Tidak Sulit
9	M9	SB	B	B	SB	B	Tidak Sulit
10	M10	B	C	B	B	B	Tidak Sulit
....
100	M100	B	C	SB	B	B	Sulit

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan sulit dan jumlah kasus keputusan tidak sulit. Menghitung *entropy* dari semua kasus. Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan ditunjukkan oleh tabel berikut ini.

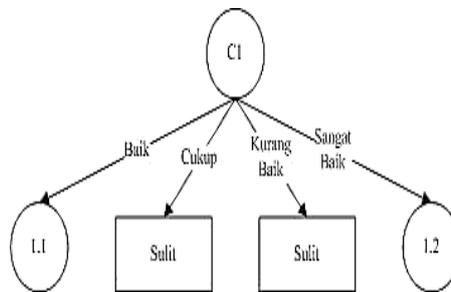
Tabel 2. Perhitungan *Node 1*

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
Total	100	52	48	0,9988455	
C1					0,23392095
SB	38	18	20	0,9980009	
B	42	14	28	0,9182958	
CB	17	17	0	0	
KB	3	3	0	0	
C2					0,14483041
SB	5	0	5	0	
B	36	12	24	0,9182958	
CB	57	38	19	0,9182958	
KB	2	2	0	0	
C3					0,181628711
SB	9	2	7	0,7642045	
B	61	24	37	0,9669853	

	CB	25	21	4	0,6343096	
	KB	5	5	0	0	
C4						0,230178262
	SB	22	7	15	0,9023933	
	B	47	17	30	0,9440870	
	CB	22	19	3	0,5746357	
	KB	9	9	0	0	
C5						0,145917982
	SB	21	6	15	0,8631206	
	B	54	25	29	0,9960384	
	CB	17	13	4	0,7871266	
	KB	8	8	0	0	

Kolom *entropy* pada baris total dihitung dengan persamaan (1) sebagai berikut :
 $Entropy (total) = - (48/100)*LOG_2(48/100) - (52/100)*LOG_2(52/100)$
 $Entropy (total) = 0,9988455$
 Kemudian pada nilai *gain* pada baris C1 dihitung persamaan (2) sebagai berikut :
 $Gain (total,C1) = (0,9988455) - ((38/100)*0,9980009 + (42/100)*0,9736681 + (17/100)*0 + (3/100)*0)$
 $Gain (total,C1) = 0,23392095$

Dari hasil perhitungan pada tabel 3 diperoleh atribut yang menjadi *node* (akar) adalah C1 memiliki gain tertinggi yaitu 0,23392095, dimana terdiri dari 4 sub atribut yaitu SB, B, CB, dan KB. Berdasarkan nilai *entropy* dari keempat sub atribut diatas, sub atribu CB dan KB yang memperoleh keputusan. Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar selanjutnya, dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 1. Pohon Keputusan 1

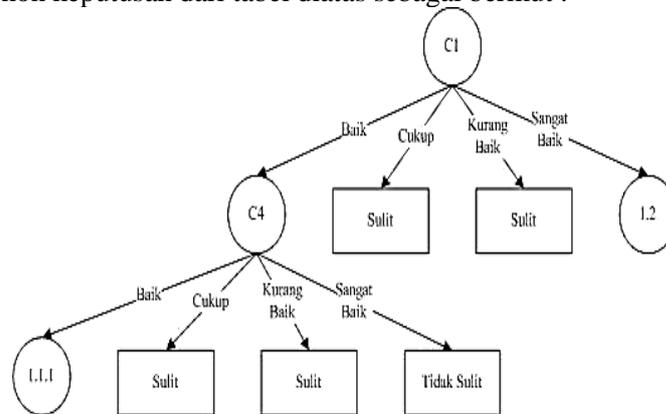
Untuk hasil perhitungan selanjutnya algoritma C4.5 dapat ditunjukkan pada tabel berikut ini

Tabel 3. Perhitungan *Node* 1.1

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	<i>Entropy</i>	<i>Gain</i>
C1 = B	42	14	28	0,9182958	
C2					0,099304428
	SB	2	0	2	0
	B	17	3	14	0,6722948
	CB	23	11	12	0,9986360
	KB	0	0	0	0
C3					0,228647718
	SB	2	0	2	0
	B	29	6	23	0,7355086
	CB	8	5	3	0,9544340
	KB	3	3	0	0
C4					0,248103743
	SB	2	0	2	0

	B	30	6	24	0,7219281	
	CB	8	6	2	0,8112781	
	KB	2	2	0	0	
C5						0,207583623
	SB	5	0	5	0	
	B	31	12	19	0,9629004	
	CB	6	2	4	0	
	KB	0	0	0	0	

Dari hasil perhitungan diatas pada tabel 4. atribut yang menjadi *Node* cabang dari atribut C1 – B adalah C4. Dimana atribut dari C4 yang terdiri dari SB, B, CB, dan KB. Dimana SB sudah memperoleh keputusan tidak sulit, CB sudah memperoleh keputusan sulit dan KB sudah memperoleh kputusan sulit. Sedangkan untuk sub atribut B belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 2. Pohon Keputusan 2

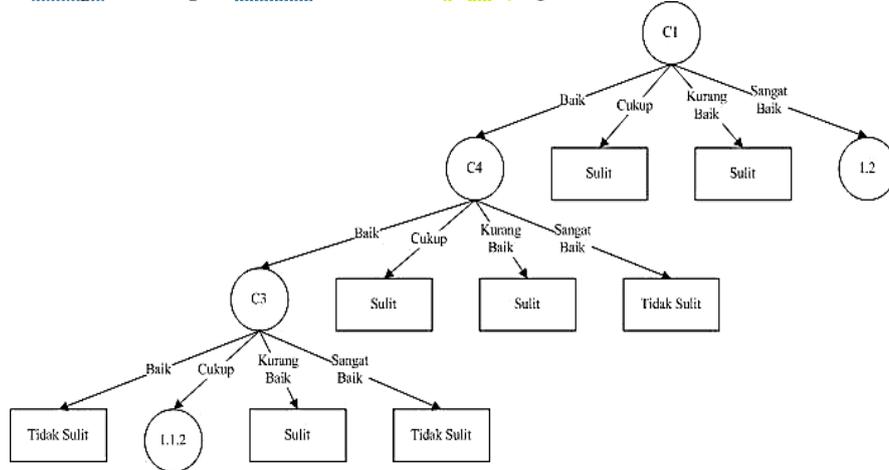
Dalam mencari hasil perhitungan seanjutnya pada *node* akar C1-B dan *node* cabang C4-B dapat ditunjukkan pada tabel berikut ini.

Tabel 4. Perhitungan *node* 1.1.1

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
C1 = B - C4 = B	30	6	24	0,7219281	
C2					0,236452798
SB	1	0	1	0	
B	14	0	14	0	
CB	15	6	9	0,9709506	
KB	0	0	0	0	
C3					0,324207948
SB	1	0	1	0	
B	24	2	22	0,4138169	
CB	2	1	1	1	
KB	3	3	0	0	
C5					0,142806945
SB	3	0	3	0	
B	23	5	18	0,7553754	
CB	4	1	3	0	
KB	0	0	0	0	

Dari hasil perhitungan pada tabel 5, atribut yang menjadi *node* cabang dari C1 – B dan C4 - B adalah C3. Dimana atribut dari C3 yang terdiri dari SB, B, CB, dan KB. Dimana SB sudah memperoleh keputusan tidak sulit, B sudah memperoleh keputusan tidak sulit

dan KB sudah memperoleh keputusan sulit. Sedangkan untuk sub atribut C belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



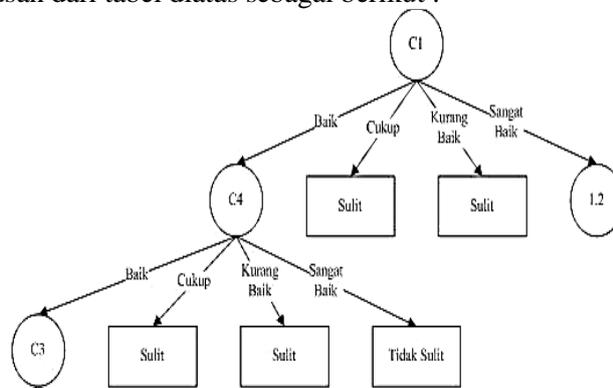
Gambar 3. Pohon Keputusan 3

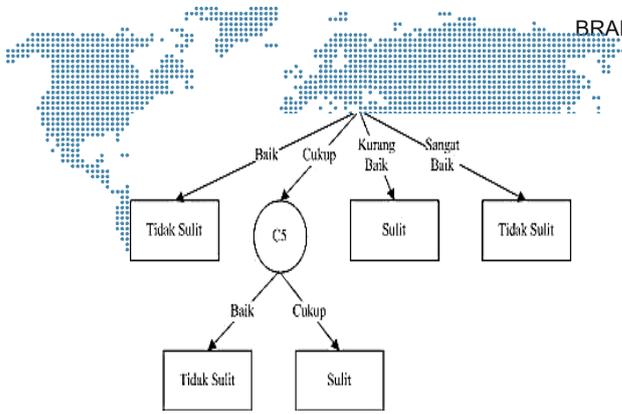
Selanjutnya melakukan perhitungan pada *node* cabang dari atribut C1-SB dapat ditunjukkan pada tabel berikut.

Tabel 5. Perhitungan *Node* 1.1.2

$C1 = B - C4 =$ $B - C3 = CB$	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
C2					0
SB	0	0	0	0	
B	0	0	0	0	
CB	2	1	1	1	
KB	0	0	0	0	
C5					1
SB	0	0	0	0	
B	1	0	1	0	
CB	1	1	0	0	
KB	0	0	0	0	

Dari hasil perhitungan diatas pada tabel 6, atribut yang menjadi *Node* cabang dari atribut $C1 = B - C4 = B - C3 = CB$ adalah C5. Dimana atribut dari C5 yang terdiri dari SB, B, CB, dan KB sudah memiliki keputusan sehingga perhitungan selesai. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :





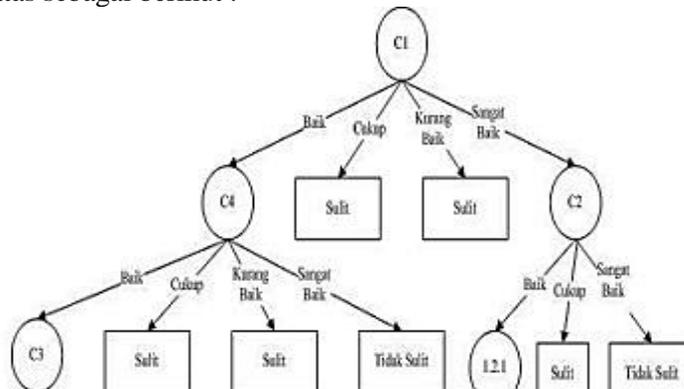
Gambar 4. Pohon Keputusan 4

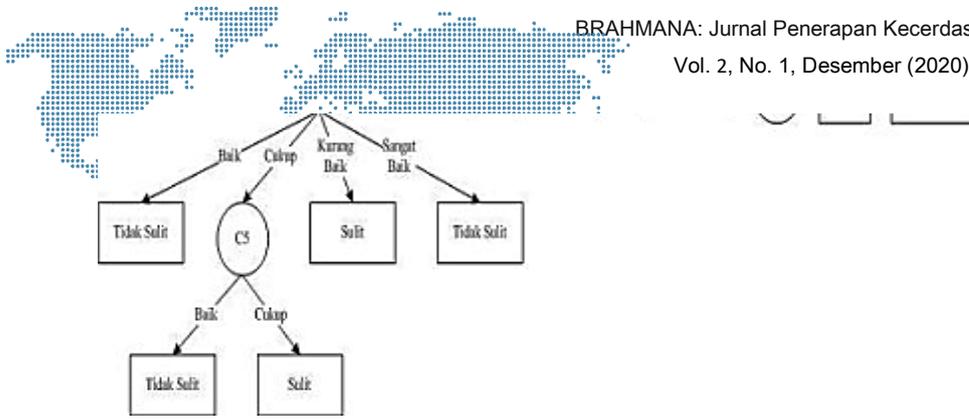
Selanjutnya melakukan perhitungan pada *node* cabang dari atribut C1-SB dapat ditunjukkan pada tabel 6.

Tabel 6. Perhitungan *Node* 1.2.

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
C1 = SB	38	18	20	0,9980009	
C2					0,14390092
SB	3	0	3	0	
B	15	5	10	0,9182958	
CB	20	13	7	0,9340681	
KB	0	0	0	0	
C3					0,043784205
SB	7	2	5	0,8631206	
B	27	13	14	0,9990103	
CB	4	3	1	0,8112781	
KB	0	0	0	0	
C4					0,137965065
SB	20	7	13	0,9340681	
B	12	6	6	1	
CB	2	1	1	1	
KB	4	4	0	0	
C5					0,069818146
SB	16	6	10	0,9544340	
B	20	10	10	1	
CB	0	0	0	0	
KB	2	2	0	0	

Dari hasil perhitungan pada tabel 7, atribut yang menjadi *node* cabang dari C1-SB adalah C2. Dimana atribut dari C2 yang terdiri dari SB, B, CB, dan KB. Dimana SB sudah memperoleh keputusan tidak sulit, CB sudah memperoleh keputusan tsulit dan KB sudah memperoleh kputusan sulit. Sedangkan untuk sub atribut B belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :





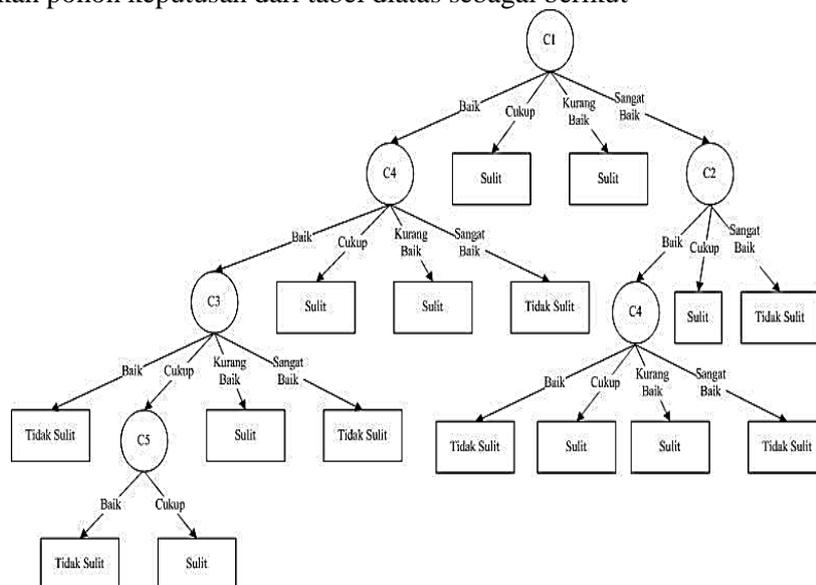
Gambar 5. Pohon Keputusan 5

Selanjutnya melakukan perhitungan pada *node* cabang dari atribut C1-SB dapat ditunjukkan pada tabel 7.

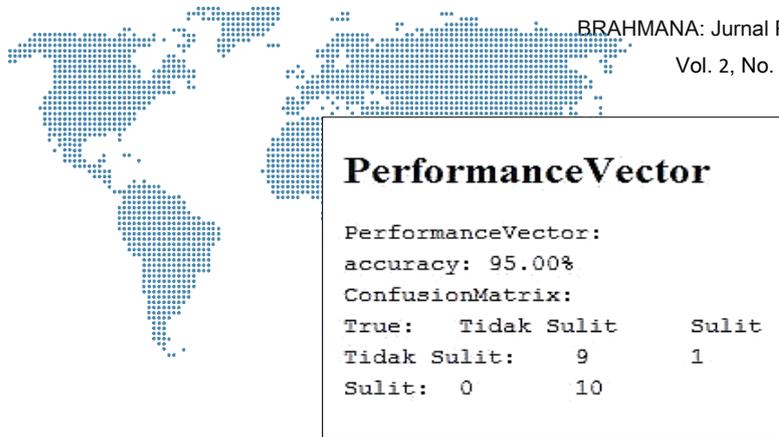
Tabel 7. Perhitungan *Node* 1.2.1

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
C1 = SB - C2 = B	15	5	10	0,9182958	
C3					0,136122932
SB	5	2	3	0,9709506	
B	9	2	7	0,7642045	
CB	1	1	0	0	
KB	0	0	0	0	
C4					0,918295834
SB	7	0	7	0	
B	3	0	3	0	
CB	1	1	0	0	
KB	4	4	0	0	
C5					0
SB	9	3	6	0,9182958	
B	6	2	4	0,9182958	
CB	0	0	0	0	
KB	0	0	0	0	

Dari hasil perhitungan pada tabel 4.6, atribut yang menjadi *node* cabang dari C1-SB dan C2-B adalah C4. Dimana atribut dari C4 yang terdiri dari SB, B, CB dan KB semua atribut sudah memperoleh keputusan sehingga perhitungan sudah selesai. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut



Gambar 6. Pohon Keputusan 6



Gambar 7. Nilai Performance Vector Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 95,00%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukan atau uji yang berasal dari data *training* dengan menggunakan *split validation* pada aplikasi *rapidminer 5.3* untuk menguji tingkat akurasi.

4. Kesimpulan

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada Penerapan Algoritma C4.5 pada klasifikasi faktor kesulitan mahasiswa dalam menyelesaikan skripsi dapat disimpulkan sebagai berikut.

- a) Permasalahan menentukan faktor kesulitan mahasiswa menyelesaikan skripsi dapat diselesaikan menggunakan teknik *data mining*, yaitu dengan Algoritma C4.5. Menghasilkan 14 (empat belas) *rules* dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 95,00%.
- b) Dari perhitungan dengan algoritma C4.5 maka didapatkan faktor yang paling dominan adalah motivasi lulus tepat waktu (C1) dengan nilai *gain* sebesar 0,204593531 .

Daftar Pustaka

- [1] M. C. Asmawan, "Analisis Kesulitan Mahasiswa Menyelesaikan Skripsi Moh.," Vol. 26, No. 2, Pp. 51–57, 2016.
- [2] A. H. Nasrullah, "Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out," Vol. 10, Pp. 244–250, 2018.
- [3] D. Nababan, "Analisa Algoritma C4 . 5 Dalam Menentukan Faktor Yang Mempengaruhi Munculnya Professional Blogger," 2019.
- [4] L. Muflikah, D. E. Ratnawati, And R. R. M. Putri, "Data Mining." .
- [5] E. Elisa, "Analisa Dan Penerapan Algoritma C4 . 5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi Pt . Arupadhatu Adisesanti," Vol. 2, No. 1, Pp. 36–41, 2017.
- [6] Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan," Vol. 12, No. 2, Pp. 116–120, 2016.
- [7] R. Wajhillah And I. Yulianti, "Penerapan Algoritma C4.5 Untuk Prediksi Penggunaan Jenis Kontrasepsi Berbasis Web," Vol. 04, No. 02, Pp. 174–187, 2017.
- [8] Muhammad, A. P. Windarto, And Suhada, "Penerapan Algoritma C4 . 5 Pada Klasifikasi Potensi," Vol. 3, Pp. 1–8, 2019, Doi: 10.30865/Komik.V3i1.1688.