

Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative  
Media 2023

## Analisis penghambat mahasiswa membuat skripsi menggunakan algoritma C4.5

Yolanda Puja Aprilia<sup>1</sup>, Agung Sutikno<sup>2</sup>

<sup>1,2</sup>Sistem Informasi, Teknik dan Ilmu Komputer, Universitas Nahdlatul Ulama Indonesia  
Jakarta 10320, Indonesia

<sup>1</sup>[yolandapuja08@gmail.com](mailto:yolandapuja08@gmail.com)

<sup>2</sup>[sagung@unusia.ac.id](mailto:sagung@unusia.ac.id)

Received on 30-10-2023, revised on 08-11-2023, accepted on 15-11-2023

### Abstrak

Penulisan skripsi merupakan bagian integral dari pendidikan tinggi yang sering kali dihadapi dengan berbagai kesulitan oleh mahasiswa. Pada penelitian ini digunakan metode algoritma C4.5 untuk mengukur dan mengidentifikasi faktor yang berpengaruh terhadap terhambatnya mahasiswa dalam menyelesaikan skripsi. Tujuan dari metode ini adalah menghasilkan *decision tree* yang dapat digunakan untuk membuat prediksi atau mengklasifikasikan data berdasarkan fitur yang ada dalam data. Dataset yang digunakan sebanyak 7 variabel diantaranya motivasi lulus tepat waktu, kemampuan menulis karya ilmiah, ketersediaan sumber belajar, proses bimbingan, lingkungan teman sebaya, tidak aktif organisasi pada semester akhir dan tanggung jawab diri sendiri selama 1 minggu dengan total dataset sebanyak 100. Pada penelitian ini menghasilkan *performance* dengan tingkat akurasi sebesar 82.00% dengan nilai *precision* sebesar 82.55% dan *recall* sebesar 84.33%. Nilai tersebut dapat membuktikan bahwa algoritma C4.5 terbukti dapat menganalisis hambatan mahasiswa dengan baik.

**Kata kunci:** mahasiswa, skripsi, algoritma C4.5, *decision tree*

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



---

#### Corresponding Author:

Yolanda Puja Aprilia  
Sistem Informasi, Teknik dan Ilmu Komputer, Universitas Nahdlatul Ulama Indonesia  
Jakarta, Indonesia  
Email: [yolandapuja08@gmail.com](mailto:yolandapuja08@gmail.com)

---

### I. PENDAHULUAN

Pendidikan memiliki peran paling utama dalam kehidupan individu, yang mengindikasikan bahwa setiap warga Indonesia berhak mendapatkan kesempatan ini dan diharapkan senantiasa mengalami perkembangan di dalamnya. Peran pendidikan memiliki dampak yang signifikan dalam persiapan dan pengembangan Sumber Daya Manusia (SDM) yang kompeten dapat berkompetisi dengan adil dan semakin meningkatnya persatuan terhadap sesama. Pendidikan merupakan dalam kategori disiplin pengetahuan yang bersifat praktis karena fokusnya adalah pada praktik dan tindakan yang mempengaruhi perkembangan para anak didik. Pembelajaran sejati diperoleh melalui pengetahuan yang dipetik dari pengalaman hidup. Oleh karena itu, pemerintah mendukung promosi pendidikan yang efektif sejak awal masa perkembangan, sehingga ketika mencapai usia dewasa akan ada Sumber Daya Manusia yang memiliki kualitas baik [1].

Seseorang yang sedang menempuh pendidikan di perguruan tinggi disebut mahasiswa [2]. Mahasiswa merupakan insan masa depan yang harus dibentuk karakternya. Proses pembentukan karakter akan terlihat jelas ketika telah menyelesaikan pendidikan di perguruan tinggi [3]. Selama proses pendidikan berlangsung, banyak tantangan yang dihadapi mahasiswa, seperti ketidakpatuhan terhadap aturan, ketidaknyamanan dalam belajar, kurangnya keseriusan dalam

menuntut ilmu, ketidaksesuaian bakat dan kemampuan dengan bidang studi, kesenjangan sosial dan keterbatasan keuangan. Semua hal ini menyebabkan ketidakkonsistenan dalam proses belajar, yang berdampak kompleks, antara lain kurangnya kesiapan dalam memahami materi, tingginya jumlah mahasiswa yang mengulang, tingkat drop out yang tinggi, serta banyak mahasiswa yang berhenti sebelum berhasil menyelesaikan pendidikan mereka.

Sebagian besar mahasiswa menghadapi kendala saat menyelesaikan Tugas Akhir (TA) atau yang lebih dikenal dengan sebutan skripsi [4]. Penyusunan skripsi merupakan tahap akhir dalam perjalanan pendidikan bagi para mahasiswa. Skripsi ini ditulis sebagai persyaratan untuk meraih kelulusan serta mendapatkan gelar sarjana yang sesuai dengan bidang studi yang diambil [2]. Faktor yang mendasari mahasiswa tidak dapat lulus tepat waktu adalah berasal dari diri sendiri dan dari lingkungan luar. Selama perjalanan pendidikan, diharapkan bahwa mahasiswa mengembangkan kemampuan dasar pada tingkat pengetahuan, keterampilan, dan sikap. Oleh karena itu, peran mahasiswa dalam konteks pendidikan tidak dapat diabaikan dari sistem pendidikan itu sendiri. Dengan demikian, mahasiswa dianggap sebagai fokus utama dari semua upaya pendidikan.

Universitas Nahdlatul Ulama Indonesia (UNUSIA) merupakan perguruan tinggi swasta Islam dilingkungan Nahdlatul Ulama. Terdapat enam Fakultas dan empat belas Program Studi yang menyelenggarakan program pendidikan dari jenjang sarjana, pascasarjana dan doktoral [5]. Penulis melakukan penelitian untuk memprediksi kelulusan mahasiswa sebagai dasar pertimbangan dan acuan dalam mengambil kebijakan akademik, dengan tujuan mengurangi tingkat putus kuliah, drop out, dan pengulangan pada mahasiswa. Penelitian ini menggunakan data faktual atau data asli. Dalam konteks penelitian ini, penulis menggunakan metode Decision Tree untuk memprediksi variabel apa saja yang berhubungan dengan kelulusan mahasiswa dan menghitung tingkat akurasi dari prediksi tersebut.

## II. METODE PENELITIAN

### A. Pengumpulan Data

Penelitian ini dilakukan dengan mengumpulkan data melalui penyebaran kuesioner dengan memberi serangkaian pertanyaan atau pernyataan tertulis kepada 100 responden yaitu mahasiswa semester 8 dan alumni Universitas Nahdlatul Ulama Indonesia. Data yang digunakan yaitu:

1. Motivasi lulus tepat waktu (X1)
2. Kemampuan menulis karya ilmiah (X2)
3. Ketersediaan sumber belajar (X3)
4. Proses bimbingan (X4)
5. Lingkungan teman sebaya (X5)
6. Tidak aktif organisasi pada semester akhir (X6)
7. Rasa tanggung jawab diri sendiri (X7)

### B. Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma Decision Tree. Algoritma ini mempunyai *input* berupa *training samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah tree. Sedangkan *samples* merupakan *field-field* data yang nantinya akan digunakan sebagai parameter. Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan [6].

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 adalah sebagai berikut [7]:

1. Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dikelompokkan kedalam kelas tertentu.
2. Setelah itu tentukan akar dari pohon. Pilih akar dari atribut, caranya adalah dengan menghitung nilai gain dari semua atribut. Atribut yang menjadi akar pertama yaitu memiliki nilai gain tertinggi. Sebelum menghitung nilai *gain*, langkah awalnya adalah menghitung nilai *entropy*.

Gunakan rumus berikut untuk menghitung nilai *entropy* [7]:

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

Keterangan:

S = himpunan kasus

N = jumlah partisi S

Pi = proporsi Si terhadap S

3. Setelah itu tentukan nilai gain menggunakan rumus [7]:

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

Keterangan:

S = himpunan Kasus

A = atribut

N = jumlah partisi atribut A

|Si| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

4. Setelah itu ulangi langkah ke-2 sampai semua *record* terpartisi secara sempurna.

5. Proses partisi pohon keputusan akan berhenti saat:

- Semua *record* dalam simpul N mendapat kelas yang sama
- Tidak ada atribut di dalam record yang dipartisi lagi
- Tidak ada record di dalam cabang yang kosong

### C. Pengolahan Data

Data yang diperoleh akan diproses dengan menggunakan perhitungan statistik deskriptif. Dimana data dari kuesioner akan dilakukan perhitungan interval. Dengan diperoleh skor terendah 7 dan skor tertinggi 28. Maka jarak antar nilai adalah 10,5. Berikut merupakan tabel kriteria tanggapan responden.

Table I. KRITERIA TANGGAPAN RESPONDEN

Rentang Skala	Kategori
17.6 - 28	Sukses
7 - 17.5	Sulit

Setiap faktor atau variabel mencakup sejumlah pertanyaan yang diajukan kepada mahasiswa yang telah dipilih. Kuesioner yang ditetapkan menggunakan linker 4 yang terdiri dari SB (Sangat Baik) dikategorikan dengan angka 4, B (Baik) dikategorikan dengan angka 3, CB (Cukup Baik) dikategorikan dengan angka 2 dan KB (Kurang Baik) dikategorikan dengan angka 1.

Table II. DATA PENELITIAN YANG SUDAH DIOLAH

Responden	X1	X2	X3	X4	X5	X6	X7	Kategori
M1	SB	B	B	B	CB	CB	SB	Tidak Sulit
M2	SB	B	B	SB	SB	CB	CB	Tidak Sulit
M3	B	CB	B	B	CB	CB	B	Tidak Sulit
M4	SB	CB	B	CB	B	KB	KB	Sulit
M5	SB	B	CB	CB	CB	KB	CB	Sulit
M6	SB	CB	CB	SB	CB	B	SB	Tidak Sulit
M7	B	SB	SB	SB	SB	SB	B	Tidak Sulit
M8	SB	KB	CB	SB	CB	KB	CB	Sulit
M9	SB	B	SB	SB	B	KB	B	Tidak Sulit
....	....	....	....	....	....	....	....	....
M100	B	CB	CB	CB	CB	B	B	Sulit

Responden	X1	X2	X3	X4	X5	X6	X7	Kategori
Responden	X1	X2	X3	X4	X5	X6	X7	Kategori
M1	SB	B	B	B	CB	CB	SB	Tidak Sulit

### III. HASIL PENELITIAN

Hasil dari *preprocessing* yang telah dilakukan sebelumnya maka selanjutnya data akan disesuaikan dengan kategori masing-masing atribut. Data kuesioner sebelumnya berupa kode diubah menjadi angka 1 sampai 4. Setiap pendapat dari responden didalamnya mempunyai kode masing-masing. Angka tersebut yang akan diolah untuk diterapkan kedalam algoritma C4.5.

	Jumlah (s)	Sulit (Si)	Tidak Sulit (Si)	Entropy	Gain
Total	100	53	47	0.997402	
X1					0.0565010
	SB	51	23	28	0.993055
	B	36	19	17	0.997772
	CB	11	9	2	0.684038
	KB	2	2	0	0
X2					0.1017553
	SB	10	4	6	0.970951
	B	59	25	34	0.983149
	CB	26	19	7	0.840359
	KB	5	5	0	0
X3					0.3372057
	SB	10	1	9	0.468996
	B	39	9	30	0.779350
	CB	47	39	8	0.658191
	KB	4	4	0	0
X4					0.2538174
	SB	18	4	14	0.764205
	B	43	15	28	0.933025
	CB	34	29	5	0.602431
	KB	5	5	0	0
X5					<b>0.3431211</b>
	<b>SB</b>	<b>12</b>	<b>0</b>	<b>12</b>	<b>0</b>
	<b>B</b>	<b>39</b>	<b>13</b>	<b>26</b>	<b>0.918296</b>
	<b>CB</b>	<b>37</b>	<b>28</b>	<b>9</b>	<b>0.800392</b>
	<b>KB</b>	<b>12</b>	<b>12</b>	<b>0</b>	<b>0</b>
X6					0.1470167
	SB	8	0	8	0
	B	21	7	14	0.918296
	CB	37	22	15	0.974025
	KB	34	24	10	0.873981
X7					0.0651555
	SB	9	1	8	0.503258
	B	36	18	18	1.000000
	CB	38	24	14	0.949452
	KB	17	10	7	0.977418

Nilai *Entropy* pada kolom total dihitung dengan rumus (1) sebagai berikut:

$$Entropy(Total) = \left( \frac{sum(yes)}{total} * \log \log 2 \left( \frac{sum(yes)}{total} \right) \right) + \left( \frac{sum(no)}{total} * \log \log 2 \left( \frac{sum(no)}{total} \right) \right)$$

$$Entropy(Total) = \left( \frac{53}{100} * \log \log 2 \left( \frac{53}{100} \right) \right) + \left( \frac{47}{100} * \log \log 2 \left( \frac{47}{100} \right) \right) = 0,997402$$

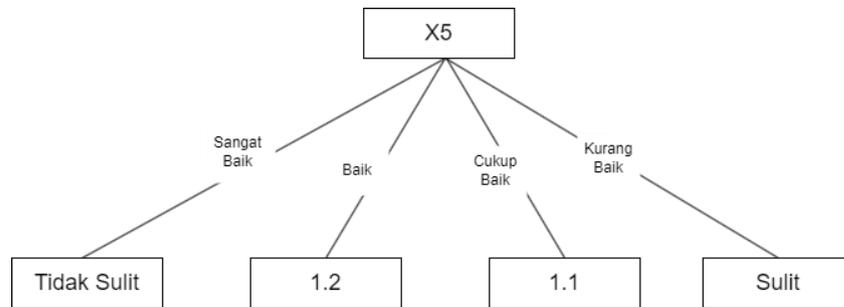
Kemudian pada nilai *gain* pada baris X5 dihitung persamaan (2) sebagai berikut:

$$Gain(total, X5) = Entropy(total) - \sum_{i=1}^n \frac{|XI|}{|Total|} * Entropy(XI)$$

$$Gain(total, X5) = 0,997402 - \left( \frac{12}{100} * 0 \right) + \left( \frac{39}{100} * 0,918296 \right) + \left( \frac{37}{100} * 0,800392 \right) + \left( \frac{12}{100} * 0 \right) = 0,3431211$$

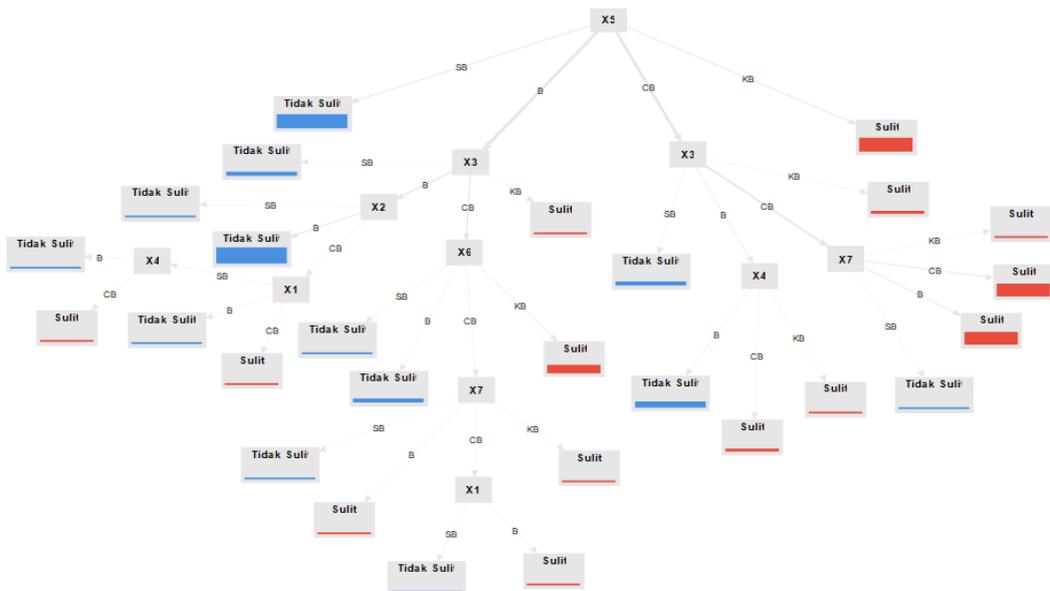
Pada tabel 3.1 bahwa atribut X5 mempunyai sub atribut SB, B, CB dan KB, dimana SB dan KB sudah memperoleh keputusan yaitu sulit dan tidak sulit. Sedangkan sub atribut B dan CB perlu dipartisi kembali.

Gambar 3.1 adalah hasil pembentukan pohon keputusan berdasarkan perhitungan yang terdapat pada tabel 2. Dari hasil perhitungan didapatkan nilai *gain* tertinggi adalah sub atribut X5, maka X5 menjadi akar pohon keputusan. Untuk menentukan akar dari atribut B dan CB, dilakukan lagi perhitungan *gain*.



Gambar 1. Pohon Keputusan

Setelah dilakukan perhitungan manual, selanjutnya mengimplementasikan sesuai *tools* yang digunakan yaitu aplikasi Rapidminer untuk menemukan keputusan sebagai pendukung dari hasil algoritma C4.5. Setelah dilakukan perhitungan dan pengujian pada data dengan menggunakan Algoritma C4.5 maka didapatkan pola keputusan akhir.



Gambar 2. Pohon Keputusan Hasil Implementasi Rapidminer

Dari pohon keputusan yang terbentuk dengan menggunakan algoritma C4.5 diperoleh *rule* berupa teks yang terbentuk dari proses penggalian data.

```

Tree
X5 = B
|
| X3 = B
| |
| | X2 = B: Tidak Sulit (Tidak Sulit=14, Sulit=0)
| | |
| | | X1 = B: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| | | X1 = CB: Sulit (Tidak Sulit=0, Sulit=1)
| | | X1 = SB
| | |
| | | X4 = B: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| | | X4 = CB: Sulit (Tidak Sulit=0, Sulit=1)
| | X2 = SB: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| | X3 = CB
| | X6 = B: Tidak Sulit (Tidak Sulit=3, Sulit=0)
| | X6 = CB
| | |
| | | X7 = B: Sulit (Tidak Sulit=0, Sulit=1)
| | | X7 = CB
| | | X1 = B: Sulit (Tidak Sulit=0, Sulit=1)
| | | X1 = SB: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| | | X7 = KB: Sulit (Tidak Sulit=0, Sulit=1)
| | | X7 = SB: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| | X6 = KB: Sulit (Tidak Sulit=0, Sulit=7)
| | X6 = SB: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| X3 = KB: Sulit (Tidak Sulit=0, Sulit=1)
| X3 = SB: Tidak Sulit (Tidak Sulit=3, Sulit=0)
X5 = CB
|
| X3 = B
| |
| | X4 = B: Tidak Sulit (Tidak Sulit=5, Sulit=0)
| | X4 = CB: Sulit (Tidak Sulit=0, Sulit=2)
| | X4 = KB: Sulit (Tidak Sulit=0, Sulit=1)
| | X3 = CB
| | X7 = B: Sulit (Tidak Sulit=0, Sulit=11)
| | X7 = CB: Sulit (Tidak Sulit=0, Sulit=11)
| | X7 = KB: Sulit (Tidak Sulit=0, Sulit=1)
| | X7 = SB: Tidak Sulit (Tidak Sulit=1, Sulit=0)
| X3 = KB: Sulit (Tidak Sulit=0, Sulit=2)
| X3 = SB: Tidak Sulit (Tidak Sulit=3, Sulit=0)
X5 = KB: Sulit (Tidak Sulit=0, Sulit=12)
X5 = SB: Tidak Sulit (Tidak Sulit=12, Sulit=0)

```

Gambar 3. Deskripsi Decision Tree

Gambar diatas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan yang telah terbentuk dengan menggunakan Algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa Algoritma C4.5 baik dalam proses penggalian data untuk menarik beberapa kesimpulan yang divisualisasikan ke dalam pohon keputusan.

Dari proses pengolahan data mahasiswa dengan menggunakan *Software Rapiminer* didapatkan hasil akurasi dengan penerapan Algoritma C4.5 sebesar 82%. Untuk lebih jelasnya dapat dilihat pada gambar berikut.

accuracy: 82.00% +/- 10.33% (micro average: 82.00%)

	true Tidak Sulit	true Sulit	class precision
pred. Tidak Sulit	37	8	82.22%
pred. Sulit	10	45	81.82%
class recall	78.72%	84.91%	

Gambar 4. Nilai Akurasi Algoritma C4.5

## PerformanceVector

```

PerformanceVector:
accuracy: 82.00% +/- 10.33% (micro average: 82.00%)
ConfusionMatrix:
True:   Tidak Sulit   Sulit
Tidak Sulit:  37      8
Sulit:   10      45
precision: 82.55% +/- 10.15% (micro average: 81.82%) (positive class: Sulit)
ConfusionMatrix:
True:   Tidak Sulit   Sulit
Tidak Sulit:  37      8
Sulit:   10      45
recall: 84.33% +/- 15.72% (micro average: 84.91%) (positive class: Sulit)
ConfusionMatrix:
True:   Tidak Sulit   Sulit
Tidak Sulit:  37      8
Sulit:   10      45
AUC (optimistic): 0.907 +/- 0.095 (micro average: 0.907) (positive class: Sulit)
AUC: 0.738 +/- 0.193 (micro average: 0.738) (positive class: Sulit)
AUC (pessimistic): 0.712 +/- 0.193 (micro average: 0.712) (positive class: Sulit)

```

Gambar 5. Nilai Performance Vektor Algoritma C4.5

Dari hasil performance vector diatas dapat diketahui bahwa penerapan Algoritma C4.5 dalam memprediksi kelulusan dengan menggunakan software RapidMiner memperoleh tingkat akurasi sebesar 82.00%. Dari gambar diatas dijelaskan bahwa prediksi sulit memiliki nilai 45 dan prediksi tidak sulit memiliki nilai 37 dengan nilai *precision* sebesar 82.55% dan *recall* sebesar 84.33%.

#### IV. PENUTUP

##### A. Kesimpulan

Berdasarkan hasil analisa data pada penelitian ini dapat diambil kesimpulan sebagai berikut:

1. Metode C4.5 dapat digunakan untuk menganalisa faktor yang menjadi kesulitan mahasiswa dalam menyusun skripsi di Universitas Nahdlatul Ulama Indonesia dengan hasil atribut Lingkungan Teman Sebaya (X5) memiliki faktor penghambat paling dominan dalam penyelesaian skripsi.
2. Penggunaan Algoritma C4.5 dalam menentukan hambatan yang dihadapi mahasiswa saat menyelesaikan skripsi memperoleh tingkat akurasi sebesar 82.00% dengan nilai *precision* sebesar 82.55% dan *recall* sebesar 84.33%. Nilai tersebut dapat membuktikan bahwa Rapidminer dapat membantu untuk membuktikan hasil perhitungan dengan metode C4.5 nilai akurasi baik.

##### B. Saran

Berdasarkan hasil yang diperoleh dari analisis dalam penelitian ini, penulis akan memberikan saran kepada peneliti yang akan datang antara lain:

1. Pada peneliti selanjutnya dapat menggunakan metode lainnya seperti *K-Nearest Neighbors* (KNN), *Naïve Bayes*, SVM untuk membandingkan tingkat ketepatan yang telah tercapai.
2. Peneliti selanjutnya juga dapat menggunakan atribut lain seperti IPK, Kehadiran untuk memprediksi faktor kesulitan mahasiswa dalam menyelesaikan skripsinya.
3. Peneliti selanjutnya dapat memperluas dataset atau peningkatan jumlah sampel sehingga dapat meningkatkan keandalan hasil penelitian.

#### V. DAFTAR PUSTAKA

- [1] N. M. A. Yayan, W. Anggraeni, U. Wiharti, "Pentingnya Pendidikan Bagi Manusia," vol. 1, no. 1, p. 67, 2019.
- [2] H. Novindari, R. Padillah, and D. Nugroho, "Pengaruh Kecemasan Mahasiswa Dan Dukungan Keluarga Terhadap Proses Penyusunan Skripsi Di Universitas Pgri Banyuwangi," *Bimbing. Dan Konseling Banyuwangi*, vol. 1, no. 1, pp. 1–6, 2022, doi: 10.36526/.Research.
- [3] M. M. Manurung and R. Rahmadi, "Identifikasi faktor-faktor pembentukan karakter mahasiswa," *JAS-PT J. Anal. Sist. Pendidik. Tinggi*, vol. 1, no. 1, p. 41, 2017.
- [4] Asmawan Moh. Chairil, "Analisis Kesulitan Mahasiswa Menyelesaikan Skripsi," *J. Penelit. Pendidik. Guru Sekol. Dasar*, vol. 26, no. Desember, p. 51, 2016.
- [5] Wikipedia, "Universitas Nahdlatul Ulama Indonesia," *Januari*, 2023. [https://id.wikipedia.org/wiki/Universitas\\_Nahdlatul\\_Ulama\\_Indonesia](https://id.wikipedia.org/wiki/Universitas_Nahdlatul_Ulama_Indonesia)
- [6] G. Lukhayu Pritalia, "Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce," *Indones. J. Inf. Syst.*, vol. 1, no. 1, pp. 47–56, 2018, doi: 10.24002/ijis.v1i1.1727.
- [7] L. Y. Lumban Gaol, M. Safii, and D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, p. 99, 2021, doi: 10.30645/brahmana.v2i2.71.