

# Prediksi Mahasiswa Berpotensi Berhenti Kuliah Secara Sepihak Menggunakan Data Mining Algoritma C4.5

Bernadus Gunawan Sudarsono<sup>1\*</sup>, Alexius Ulan Bani<sup>2</sup>

<sup>1</sup>Program Studi Sistem Informasi, Universitas Bung Karno Jakarta

<sup>2</sup>Program Studi Komputer, Universitas Bung Karno Jakarta  
gunawanbernadus@ubk.ac.id<sup>1\*</sup>

## Abstract

*Students are a call to a student who has sat in college, there are many reasons why students decide to stop taking one-sidedly lectures which result in many losses that will be experienced by the campus, starting from reduced campus quantisas and campus quality, adding to the accumulation of unclear student data and will cause and slow down the campus reporting performance, so it is necessary to predict students who have the potential to stop studying unilaterally by looking at several criteria and digging up information on potential student data by applying data mining science and approaching one of the classification approach algorithms, namely the C4.5 algorithm, The results obtained from the study are able to see the predictive rule of students who have the potential to unilaterally quit college.*

**Keywords:** *students dropping out of college, Data Mining, C4.5 Algorithm*

## Abstrak

*Mahasiswa merupakan panggilan terhadap seorang pelajar yang telah duduk dibangku perguruan tinggi, sangat banyak penyebab mahasiswa memutuskan untuk berhenti mengikuti perkuliahan secara sepihak yang mengakibatkan banyak kerugian yang akan dialami pihak kampus, mulai dari berkurangnya kuantisas kampus dan kualitas kampus, bertambahnya penumpukan data mahasiswa yang tidak jelas dan akan menyebabkan dan memperlambat kinerja pelaporan kampus, sehingga diperlukan prediksi mahasiswa yang berpotensi berhenti kuliah secara sepihak dengan melihat beberapa kriteria dan menggali informasi terhadap data mahasiswa berpotensi dengan menerapkan ilmu data mining dan melakukan pendekatan terhadap salah satu algoritma pendekatan klasifikasi yaitu algoritma C4.5, didapatkan hasil dari penelitian mampu melihat rule prediksi terhadap mahasiswa yang berpotensi berhenti kuliah secara sepihak.*

**Kata Kunci:** *mahasiswa berhenti kuliah, Data Mining, Algoritma C4.5*

## 1. PENDAHULUAN

Pendidikan merupakan gerbang dari masa depan bangsa dan negara, Pendidikan merupakan keharusan yang mutlak dimiliki oleh anak bangsa, pendidikan tidak mengenal usia, Pendidikan berfungsi untuk mendidik dalam menciptakan generasi yang terarah dalam menggapai impian, cita-cita dan turut mendidik disiplin, etika baik, dan cinta terhadap negara, Pendidikan juga merupakan cerminan dari bangsa itu sendiri. Setiap individu diwajibkan mendapatkan Pendidikan minimal hingga jenjang sekolah menengah pertama atau pun sekolah menengah kejuruan sebagai syarat sudah terdidiknya tenaga kerja yang dapat bersaing dalam dunia kerja dan dunia usaha.[1]

Bangsa yang maju merupakan bangsa yang mampu melakukan perubahan baik dalam bidang pembangunan maupun bidang ekonomi,

generasi muda yang terdidik dan memiliki pendidikan yang tinggi juga merupakan asset yang sangat berharga bagi negara untuk melanjutkan perjuangan para kaum tua yang sudah membesarkan dan membuat bangsa Indonesia menjadi salah satu negara yang terdaftar sebagai negara yang berkembang. Baik buruknya suatu bangsa dan nasib bangsa Indonesia juga ada pada generasi yang terdidik.

Pendidikan memiliki jenjang beraneka ragam mulai dari Sekolah Dasar (SD) yang ditempuh selama enam tahun dan biasanya usia anak bersekolah pada taraf 6 -7 tahun, sekolah menengah pertama (SMP) dan sederajatnya, SMK dan SMA,MA dan pada tingkat selanjutnya yaitu Mahasiswa. Mahasiswa sendiri merupakan sebutan untuk orang-orang yang memilih menempuh pendidikan diatas SMA, SMK sederajat atau sering disebut dengan diatas rata-rata siswa yang sedang mengikuti Pendidikan perguruan tinggi[2].

Untuk menempuh perguruan tinggi sangat banyak memakan waktu dan terutama biaya yang cukup besar, sangat banyak masyarakat indonesia yang memilih melanjutkan pendidikan demi mendapat gelar ilmu pengetahuan yang mereka minati dan sebagai investasi masa depan bagi anak bangsa, dalam disiplin ilmu perguruan tinggi mahasiswa dituntut agar lebih objektif dan lebih aktif dalam mencari ilmu secara mandiri dan terpadu sehingga Ketika mahasiswa lulus dari jenjang Pendidikan tinggi menunjukkan mahasiswa merupakan orang-orang yang sangat terlatih dan terdidik dalam aspek ilmu tertentu dan membuat para pengusaha, instansi pemerintahan, perusahaan besar, industri lebih memilih untuk mempekerjakan lulusan dari sebuah perguruan tinggi dan hal itu menjadi alasan utama banyak masyarakat memilih untuk melanjutkan pendidikan kejenjang lebih tinggi.

Dari banyaknya masyarakat yang memilih melanjutkan Pendidikan tinggi dan tercatat sebagai mahasiswa salah satu lembaga pendidikan, banyak juga mahasiswa yang memilih berhenti secara sepihak dan menarik diri untuk tidak mengikuti perkuliahan sama sekali dikarenakan beberapa faktor mulai dari mahasiswa tersebut mengalami masalah keluarga, masalah keuangan, masalah nilai, masalah kehadiran, bisa juga masalah pribadi antara mahasiswa dan dosen atau staf pengajar. Berhentinya mahasiswa kuliah secara sepihak akan memberikan dampak tersendiri terhadap kampus, terkadang mahasiswa yang berhenti kuliah secara sepihak akan berlangsung selama bertahun-tahun dan sangat banyak juga membuat mahasiswa harus di *drou out* atau di keluarkan dari kampus atau lebih tepatnya dari data kampus yang menyatakan sebagai mahasiswa aktif menjadi mahasiswa non aktif.

Sangat banyak mahasiswa yang sudah dinyatakan non aktif dari kampus dikarenakan masa berhenti kuliah sudah sampai dua tahun lebih, mahasiswa yang berhenti secara tiba-tiba akan menyebabkan kampus kekurangan kuantitas dan dapat mempengaruhi mahasiswa lainnya untuk ikut berhenti kuliah dan menyebabkan terjadinya troble data pada system pelaporan dan pengamatan mahasiswa khususnya akan membuat sangat sulit pihak *feeder* kampus, pihak program studi dan terjadinya penumpukan

data mahasiswa yang tidak jelas. Dan tidak jarang pula mahasiswa yang sudah berhenti kuliah dan Namanya sudah dihapuskan meminta untuk Kembali diizinkan mengikuti perkuliahan kembali sesuai dengan semester yang telah mereka jalani, hal tersebut membuat pekerjaan pihak kampus semakin banyak dan harus menjalani prosedur yang sangat Panjang, hal tersebut juga dianggap merugikan pihak kampus dan prosedur yang telah dijalani.

Untuk itu perlu dilakukan pemantauan, prediksi dan evaluasi terhadap mahasiswa yang berpotensi berhenti kuliah secara sepihak agar dapat mengurangi tingkat berhentinya mahasiswa dalam masa perkuliahannya, hal tersebut dapat dilakukan dengan menggunakan metode data mining. Data Mining berfungsi untuk mencari danmendapatkan informasi yang dibutuhkan dari berbagai tumpukan data yang banyak secara komputerisasi[3].

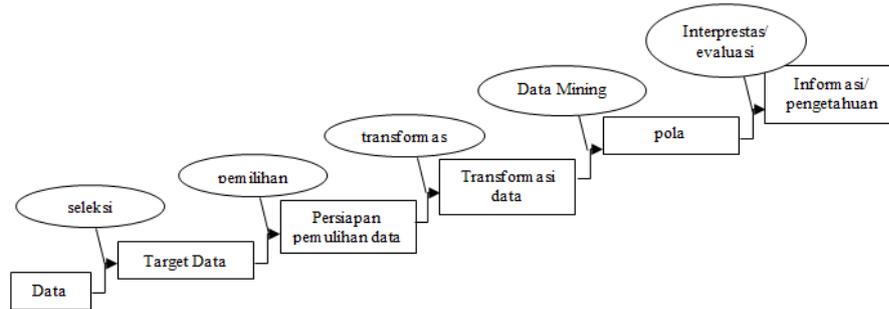
Pada penelitian terhdahulu penggunaan data mining dalam prediksi sudah sangat banyak digunakan dengan menggunakan beberapa algoritma pendekatan data mining yang berbeda, penelitian sebelumnya telah terbukti mampu dalam menyelesaikan masalah data pada prediksi penjualan pada toko disuatu daerah tertentu, data mining terhadap prediksi diagnose penyakit jantung, ginjal, kanker dan berbagai penyakit lainnya yang bertujuan agar dapat ditangani lebih dini dari informasi yang diolah menggunakan data mining[4][5]. Algoritma yang digunakan dalam prediksi suatu data untuk mendapatkan satu innformasi biasanya peneliti terdahulu seperti algoritma Apriori, Assosiatoin Rule, Rought Set, C4.0, C4.5, Naïve Bayes dan lainnya. Pada penelitian ini peneliti menggunakan algoritma C4.5 untuk melihat kecenderungan mahasiswa terhadap potensi berhenti kuliah secara sepihak dengan menggambarkan sebuah pohon keputusan atau yang lebih dikenal dengan *decision Tree* dan merupakan salah satu cabang ilmu KDD atau *knowledge Discovery in Database*[6].

## 2. METODOLOGI PENELITIAN

Pada penelitian ini menggunakan metode kuantitatif dalam proses analisa dengan menggunakan teknik klasifikasi yang nantinya menghasilkan pohon keputusan yang akan membantu dalam hal membaca sebuah situasi dan kondisi berdasarkan Analisa menggunakan algoritma C4.5 [7].

### 2.1. KDD (*Knowledge Discovery in Database*)

KDD yaitu merupakan singkatan dari *Knowledge discovery in database* yang disebutkan dengan cara menggali data dari sebuah tumpukan data untuk mendapatkan informasi yang dibutuhkan, proses ini dilakukan dengan menggunakan sistem komputer, dalam proses kdd data mining sebagai berikut [8]:



**Gambar 1.** Proses Knowledge Discovery Database (KDD)

### 2.1. Algoritma C4.5

Algoritma C4.5 merupakan salah satu dari metode klasifikasi dimana algoritma ini pada keputusan akhir menggunakan *decision tree* atau sebuah pohon keputusan dari keputusan tersebut akan dibentuk sebuah cabang yang memperlihatkan hasil dari setiap perhitungan menggunakan algoritma C4.5, pohon keputusan tersebut yang akan menunjukan beberapa hubungan antara tiap item atribut yang tersembunyi, model pohon keputusan ini biasa disebut dengan prediksi menggunakan struktur hirarki yang embentu sebuah pohon yang bukan memiliki daun tetapi dalam bentuk node. penyelesaian menggunakan algoritma ini dilakukan beberapa tahapan berupa[9][10]:

- a) Menyiapi data yang akan diuji atau biasanya sering disebut dengan penyiapan data training dari data lama yaitu data atau histori yang sudah pernah terjadi sebelumnya dengan data yang baru sebagai bahan pertimbangan dalam pengambilan keputusan dari tinjauan data yang pernah ada.
- b) Menentukan akar dari sebuah pohon keputusan yang diambil dari mencari rumus atau nilai Gain, nilai Gain dapat disimpulkan jika sudah mendapatkan nilai *Entropy* sebelumnya, *entropy* didapatkan dengan menggunakan rumusan berikut:

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

Keterangan:

S : himpunan dari sebuah kasus

n : jumlah

pi: nilai keseimbangan dari Si terhadap S

- c) Mencari nilai Gain

$$Gain (S, A) = Entropy (s) - \sum_{i=1}^n \frac{|s_i|}{|s|} * Entropy \quad (2)$$

Keterangan:

|s<sub>i</sub>| : jumlah kasus pada ke-i

|s| : jumlah kasus dalam atribut S

A : Atribut

n : jumlah atribut A

- d) Lakukan seluruh proses mulai dari pengkelompokan terhadap pemilihan atribut hasil, pencarian nilai *entropy* hingga nilai gain, hingga keseluruhan data terpastisi.
- e) Lakukan pembentukan pohon keputusan jika sudah mendapatkan nilai gain.

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini data diambil dari kampus Universitas Budi Darma terlihat dari data mahasiswa yang masih aktif dan mengikuti perkuliahan dengan beberapa mahasiswa yang melakukan berhenti secara sepihak menggunakan Analisa algoritma C4.5 dimana dikelompokan dulu setiap kriteria dengan hasil terlebih dahulu agar memudahkan dalam perhitungan nilai *entropy* dan gain untuk membuat sebuah pohon keputusan dan menghasilkan prediksi yang tepat. Berikut merupakan tahapan penyelesaian kasus.

**Tabel 1.** kasus lama data mahasiswa

No	Nama	Npm	Kelas	Nilai	Kehadiran	Pembayaran Uang Kuliah (PKU)	Status
1	Fahrul Rozi Iskandar	14110234	Sore	Sedang	Jarang Hadir	Tidak Lancar	Tidak Aktif
2	Rasyid Arman	14110026	Pagi	Rendah	Sering hadir	Lancar	Aktif
3	Bela Sari	14112120	Pagi	Sedang	Sering Hadir	Lancar	Aktif
4	Bayu	14110131	Sore	Rendah	Jarang Hadir	Lancar	Aktif
5	Fajri Nasution	14110055	Sore	Rendah	Jarang Hadir	Tidak Lancar	Tidak Aktif
6	Pintauli	14110021	Pagi	Rendah	Sering Hadir	Tidak Lancar	Tidak Aktif
7	Indri Dewi	14105611	Sore	Tinggi	Sering Hadir	Tidak Lancar	Aktif
8	Ali Rahmad Pohan	15001101	Pagi	Tinggi	Sering Hadir	Lancar	Aktif
9	Devi Afnita	15230002	Pagi	Tinggi	Sering Hadir	Lancar	Aktif

#### 3.1. Pembahasan

Pada pembahasan ini dilakukan perhitungan pengelompokan terhadap data dengan membuat tabel yang berisi node untuk pembentukan terhadap akar awal dapat dilihat pada tabel 2 berikut.

**Tabel 2.** Pembentukan Akar Awal Node 1

No		Keterangan	Jumlah kasus (s)	Aktif	Tidak Aktif	Entropy	Gain
1	Total		9	6	3	0,918296	
	Nilai						0,251629
		Rendah	4	2	2	1	
		Tinggi	3	3	0	0	
		Sedang	2	1	1	1	
	Kehadiran						0,178849
		Jarang Hadir	3	1	2	0,918296	
		Sering Hadir	6	5	1	0,650022	
	Pembayara						0.83442

No	Keterangan	Jumlah kasus (s)	Aktif	Tidak Aktif	Entropy	Gain
	Pembayaran Uang Kuliah (PKU)					
	Lancar	5	5	0	0	
	Tidak Lancar	4	1	3	0,188721	

a) Mencari Nilai *Entropy*

$$\text{Total} = \left(-\frac{6}{9} * \log_2 \frac{6}{9}\right) + \left(-\frac{3}{9} * \log_2 \frac{3}{9}\right) = 0,918296$$

$$\text{nilai rendah} = \left(-\frac{2}{4} * \log_2 \frac{2}{4}\right) + \left(-\frac{2}{4} * \log_2 \frac{2}{4}\right) = 1$$

$$\text{nilai Tinggi} = \left(-\frac{3}{3} * \log_2 \frac{3}{3}\right) + \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) = 0$$

$$\text{nilai Sedang} = \left(-\frac{1}{2} * \log_2 \frac{1}{2}\right) + \left(-\frac{1}{2} * \log_2 \frac{1}{2}\right) = 1$$

$$\text{kehadiran, jarang hadir} = \left(-\frac{1}{3} * \log_2 \frac{1}{3}\right) + \left(-\frac{2}{3} * \log_2 \frac{2}{3}\right) = 0,918296$$

$$\text{kehadiran, Sering hadir} = \left(-\frac{5}{6} * \log_2 \frac{5}{6}\right) + \left(-\frac{1}{6} * \log_2 \frac{1}{6}\right) = 0,65002$$

$$\text{PKU Lancar} = \left(-\frac{5}{5} * \log_2 \frac{5}{5}\right) + \left(-\frac{0}{5} * \log_2 \frac{0}{5}\right) = 0$$

$$\text{PKU Tidak Lancar} = \left(-\frac{1}{4} * \log_2 \frac{1}{4}\right) + \left(-\frac{3}{4} * \log_2 \frac{3}{4}\right) = 0,188721$$

b) Mencari Nilai Gain

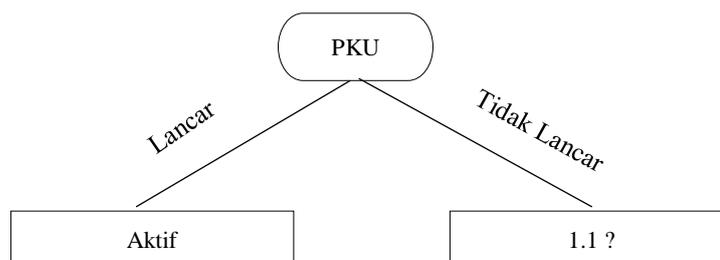
$$\text{Nilai} = 0,918296 - \left(\left(\frac{4}{9} * 1\right) + \left(\frac{3}{9} * 0\right) + \left(\frac{2}{9} * 1\right)\right) = 0,251629$$

$$\text{Kehadiran} = 0,918296 - \left(\left(\frac{3}{9} * 0,918296\right) + \left(\frac{6}{9} * 0,65002\right)\right) = 0,178849$$

$$\text{Pembayaran Uang Kuliah} = 0,918296 - \left(\left(\frac{5}{9} * 0\right) + \left(\frac{4}{9} * 0,188721\right)\right) = 0,83442$$

c) Pohon Keputusan

Pohon keputusan awal dibuat berdasarkan nilai gain tertinggi. Nilai gain tertinggi ada pada tabel milik oleh pembayaran uang kuliah, sehingga pohon keputusan dapat dibentuk sebagai berikut ini:



**Gambar 2.** Pohon keputusan

d) Lakukan perhitungan Kembali jika masing-masing node dari cabang tidak memiliki nilai keputusan atau *entropy* tidak bernilai 0 dari data pencarian nilai tertinggi terhadap gain.

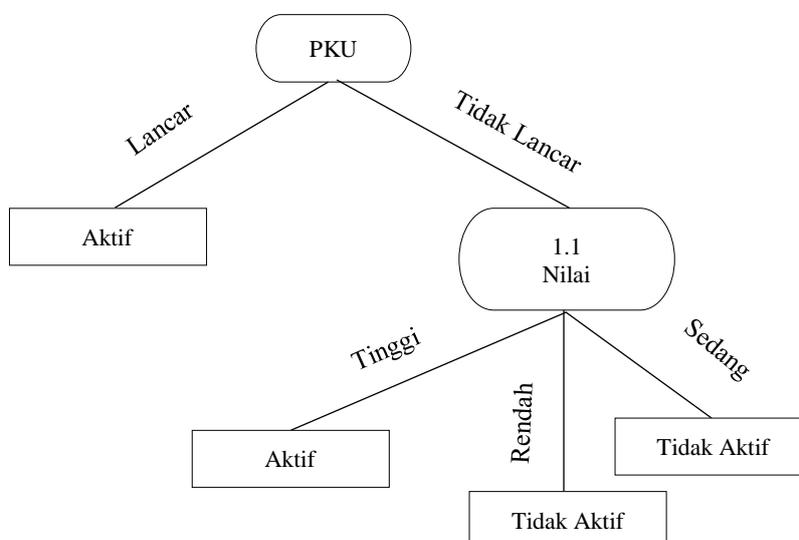
**Tabel 3.** kasus yang belum mendapatkan nilai *entropy*

No	Nama	Npm	Kelas	Nilai	Kehadiran	PKU	Status
1	Fahrul Rozi Iskandar	14110234	Sore	Sedang	Jarang Hadir	Tidak Lancar	Tidak Aktif
2	Fajri Nasution	14110055	Sore	Rendah	Jarang Hadir	Tidak Lancar	Tidak Aktif
3	Pintauli	14110021	Pagi	Rendah	Sering Hadir	Tidak Lancar	Tidak Aktif
4	Indri Dewi	14105611	Sore	Tinggi	Sering Hadir	Tidak Lancar	Aktif

**Tabel 4.** Pembentukan Akar Awal Node 1

No		Keterangan	Jumlah kasus (s)	Aktif	Tidak Aktif	<i>Entropy</i>	Gain
1	Pembayaran Uang Kuliah (PKU)-	Tidak Lancar	4	1	3	0,811278	
	Nilai						0,811278
		Rendah	2	0	2	0	
		Tinggi	1	1	0	0	
		Sedang	1	0	1	0	
	Kehadiran						0,764098
		Jarang Hadir	2	0	2	0	
		Sering Hadir	2	1	1	1	

Dari hasil nilai Gain tertinggi terletak pada atribut Nilai yang membentuk pohon keputusan sebagai berikut ini:



**Gambar 3.** Pohon keputusan Node 1.1

#### 4. KESIMPULAN

Pada pernyataan diatas menjelaskan bahwa hasil keputusan terhadap prediksi tertinggi yang mempengaruhi berhentinya mahasiswa secara sepihak secara sengaja maupun tidak terdapat penyebab tertinggi berhentinya mahasiswa kuliah secara sepihak dikarenakan Pembayaran Uang kuliah yang Tidak Lancar dan memiliki nilai yang Rendah, peluang lainnya ada pada nilai yang tergolong biasa saja atau sedang tetap terpengaruh pada ketidak mampuan membayar uang kuliah.

Untuk mengurangi resiko diatas pihak kampus sudah bisa melakukan evaluasi terhadap penyebab terjadinya berkurangnya jumlah mahasiswa secara signifikan setiap tahunnya atau mahasiswa yang melakukan berhenti kuliah secara sepihak dengan cara melakukan pendekatan dan support terhadap mahasiswa dengan kebijakan manajemen dan startegi dan jaminan mutu perguruan tinggi untuk lebih mengatur terhadap permasalahan tersebut untuk menghindari hal yang tidak di inginkan.

#### DAFTAR PUSTAKA

- [1] Sarnawi M Dasim, "Implementasi Pendidikan Karakter Dalam Pembelajaran Sains Di Sekolah Dasar Universitas Pendidikan Indonesia," *Thesis*, pp. 78–95, 2012.
- [2] N. Ilmianah and P. Karakter, "Penguatan pendidikan karakter di smk dalam era mea," pp. 374–387, 2003.
- [3] D. Evanko, "Optical imaging of the native brain," *Nat. Methods*, vol. 7, no. 1, p. 34, 2010, doi: 10.1038/nmeth.f.284.
- [4] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," *Data Min. Concepts Tech.*, 2012, doi: 10.1016/C2009-0-61819-5.
- [5] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [6] D. P. Utomo, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," vol. 4, no. April, pp. 437–444, 2020, doi: 10.30865/mib.v4i2.2080.
- [7] A. Nugroho, *Analisis dan Perancangan Sistem Informasi dengan Metodologi Berorientasi Objek*, Edisi Revi. Bandung: Informatika Bandung, 2005.
- [8] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [9] R. Rahim, E. Buulolo, and N. Silalahi, "C4.5 Algorithm To Predict The Impact Of The Earthquake," no. January, pp. 2–8, 2017, doi: 10.31227/osf.io/rbwmg.
- [10] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining Untuk

Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Issn 2548-8368*, vol. Vol 1, No, no. June, p. 7, 2017, doi: 10.30865/mib.v1i2.323.

- [11] M. Muhamad, A. P. Windarto, and S. Suhada, "Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 1-8, 2019, doi: 10.30865/komik.v3i1.1688.