

Penerapan Data Mining dengan Algoritma C4.5 Dalam Memprediksi Penjualan Tempe

Yusuf Maulana^{1*}, Riki Winanjaya¹, Fitri Rizki²

¹STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

²AMIK Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹Koesmadji46@gmail.com,

Abstrak—Banyak orang membicarakan tentang bisnis, keuntungan bisnis serta banyaknya tipe bisnis yang berbagai macam yang pasti mempunyai satu tujuan yaitu memperoleh keuntungan secara finansial ataupun laba. Dengan begitu banyak jenis bisnis serta khasiat yang dapat diambil dari bisnis pasti kerap kali kita merasa ingin lekas terjun ke dunia bisnis buat memperoleh keuntungan ataupun lebih spesifik ialah pemasukan. Salah satu contoh bisnis ialah berjualan tempe. Penjualan merupakan aktivitas menjual sesuatu produk ataupun jasa yang butuh dicermati oleh pelakon bisnis maupun usaha semacam usaha penjualan tempe pada ac mandiri. Didalam berjualan tentu kita memiliki sekumpulan informasi penjualan yang sangat banyak sebab itu kita wajib bisa mencerna informasi penjualan supaya jadi data baru. Dalam permasalahan ini algoritma c4.5 ialah suatu tata cara yang dapat membantu mencerna ataupun memprediksi nilai penjualan pada waktu yang hendak tiba. Riset ini dicoba dalam rangka menolong para penjual buat memprediksi penjualan dagangannya sehingga mereka bisa mempersiapkan ataupun menyetok bahan yang diprediksi akan menghadapi peningkatan dalam penjualan nya sekaligus mencerna informasi supaya informasi tersebut bisa berguna serta menghasilkan suatu data yang baru ialah rapid miner. Dari riset ini bisa disimpulkan kalau dengan memakai algoritma c4.5 penjualan tempe bisa diprediksi dengan akurasi yang cukup tinggi.

Kata Kunci: Data Mining; Penjualan; Algoritma C4.5; Rapid Miner; Prediksi

Abstract—Many people talk about business, benefits of business and the many different types of businesses that definitely have one purpose, financial profit or profit. With so many types of business as well as the benefits that can be taken from business, often we definitely feel we feel quickly to pave in the world of business to get profit or more specifically is income. One example of business is selling tempe. Sales is the activity of selling a product or service that needs business authorities or business acknowledgments of a kind of business selling tempe at mandiri ac. In selling of course we have a very many collection of sales information why we should be able to digest the sales information to become new data. In this issue the c4.5 algorithm is a procedure that can help digest or predicate the value of sales at the time to arrive. This research was tested in order to help sellers to predict the sales of their merchantability so that they can prepare or stock materials which is predicted to face an increase in their sales at the time of getting the information between supplied information. From this research can be conclusioned if using the c4.5 algorithm the sales of tempe can be predicted with a quite high accuracy.

Keywords: Data Mining; Sales; C4.5 Algorithm; Rapid Miner; Prediction

1. PENDAHULUAN

Tempe merupakan makanan yang diolah dari fermentasi kacang kedelai yang bergizi dan sangat digemari oleh penduduk Indonesia sebagai bahan pelengkap lauk pauk. Tempe juga sangat terkenal di berbagai negara eropa selain itu tempe mempunyai zat antibakteri untuk mencegah penyakit jantung dan juga penyakit kolestrol. Selain memiliki nilai gizi yang tinggi, harga tempe yang murah menjadikan minat masyarakat untuk mengonsumsi tempe relatif tinggi. Tempe sendiri merupakan bahan makanan yang mudah diolah menjadi berbagai macam olahan makanan. Tingginya minat masyarakat dalam mengonsumsi tempe menjadikan banyak peluang usaha. Salah satu perusahaan yang bergerak dibidang penjualan tempe adalah AC Mandiri. AC Mandiri berlokasi di Jalan Rakutta Sembiring Pematangsiantar. Namun, hasil penjualan yang bervariasi menjadikan perusahaan ini harus mengatur strategi agar terhindar dari kerugian mengingat banyaknya pesaing. Berdasarkan masalah tersebut, klasifikasi dapat dimanfaatkan menjadi salah satu rekomendasi pemecahan masalah untuk memprediksi penjualan tempe pada AC Mandiri. Salah satu metode pemecahan masalah yang digunakan dalam penelitian adalah Data Mining sebagai rekomendasi solusi dari permasalahan. Data mining sering disebut *knowledge discovery in database* (KDD), kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteratur atau hubungan dalam set data berukuran besar. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis* [1]. Salah satu algoritma yang termasuk dalam datamining klasifikasi yaitu C4.5. Algoritma C4.5 yaitu sebuah Algoritma yang digunakan untuk membangun decision tree pengambilan keputusan untuk memilih atribut akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Model yang di peroleh bisa berbentuk aturan IF-THEN, pohon keputusan, rumus matematika atau jaringan saraf tiruan dengan cara mengetahui kasus yang ada dapat diekspetasi dari pohon keputusan.

Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik [2]. Penerapan algoritma Data Mining C4.5 dalam sistem klasifikasi telah banyak dilakukan dalam penelitian terdahulu. Salah satu diantaranya yaitu penerapan metode klasifikasi *Decision Tree* dan algoritma C4.5 dalam Memprediksi Kriteria Nasabah Kredit Mega *Auto Finance* [3]. Salah satu bentuk nyata dari upaya ini ialah menghitung kemampuan penjualan seakurat mungkin sehingga dapat diketahui seberapa banyak penjualan pada AC Mandiri. Algoritma yang hendak digunakan buat memprediksi penjualan dalam permasalahan ini merupakan algoritma C4. 5 tata

cara ini hendak berupaya memperoleh prediksi penjualan serta bisa diketahui turun naiknya penjualan pada industri tersebut.

2. METODOLOGI PENELITIAN

2.1 Proses Pengumpulan data

Penelitian ini dilakukan di AC Mandiri yang beralamat di Jalan Rakutta Sembiring di kota pematang siantar dengan menghabiskan waktu dari tanggal 25 April sampai 03 Mei. Sumber data diperoleh dari dokumen dokumen tertulis yang terdata di AC Mandiri.

2.2 Analisa Data

Analisis data merupakan sebuah proses yang dilakukan untuk mengumpulkan data yang sudah selesai. Dalam riset ini peneliti sudah mendapatkan data sehingga dapat langsung digunakan untuk data penelitian. Data yang dipakai merupakan data primer. *Data primer* ialah data yang mempunyai sifat *up to date* dan untuk mendapatkan sebuah *data primier* peneliti harus terjun langsung ke lapangan untuk wawancara, observasi dan diskusi dengan pemilik AC Mandiri. Adapun *variable* yang digunakan untuk memprediksi penjualan tempe pada AC Mandiri yaitu : (1) permintaan: naik dan turun, (2) persediaan : banyak dan sedikit (3) produksi : bertambah dan berkurang, dapat dilihat pada tabel 1 berikut.

Tabel 1. Data Penjualan Tempe Selama Dua Tahun

2019/2020	Jumlah	Harga/Kg	Penjualan	Kemasan	Status
Januari	618 kg	16.000	Produsen	Daun	Tinggi
Februari	712 kg	16.000	Pedagang perantara	Daun	Rendah
Maret	745 kg	16.000	Produsen	Plastik	Rendah
April	685 kg	16.000	Produsen	Daun	Rendah
Mei	815 kg	15.000	Pedagang perantara	Plastik	Tinggi
Juni	745 kg	16.000	Pedagang perantara	Plastik	Rendah
Juli	865 kg	16.000	Produsen	Daun	Tinggi
Agustus	882 kg	15.000	Pedagang perantara	Daun	Tinggi
September	845 kg	15.000	Pedagang perantara	Plastik	Tinggi
Oktober	692 kg	16.000	Pedagang perantara	Plastik	Rendah
November	855 kg	15.000	Pedagang perantara	Daun	Tinggi
Desember	876 kg	15.000	Produsen	Daun	Tinggi
Januari	885 kg	16.000	Produsen	Daun	Tinggi
Februari	715 kg	16.000	Pedagang perantara	Daun	Rendah
Maret	734 kg	16.000	Produsen	Plastik	Tinggi
April	767 kg	15.000	Produsen	Daun	Tinggi
Mei	789 kg	15.000	Pedagang perantara	Plastik	Rendah
Juni	795 kg	15.000	Produsen	Plastik	Tinggi
Juli	743 kg	15.000	Pedagang perantara	Daun	Tinggi
Agustus	875 kg	15.000	Pedagang perantara	Daun	Tinggi
September	898 kg	15.000	Pedagang perantara	Daun	Tinggi
Oktober	895 kg	15.000	Produsen	Plastik	Rendah
November	725 kg	17.000	Pedagang perantara	Plastik	Rendah
Desember	946 kg	17.000	Produsen	Daun	Tinggi

2.3 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang dapat digunakan dalam membentuk suatu pohon keputusan, dimana metode pohon keputusan mengubah fakta yang besar menjadi sebuah pohon keputusan yang dapat mempresentasikan aturan [4]–[10]. Adapun cara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

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

Hal pertama yang dilakukan untuk memilih atribut sebagai akar diperlukan nilai tertinggi pada atribut-atribut yang ada, dan rumus yang dipakai untuk dasar entropi adalah sebagai berikut:

$$Entropi (S) = - \sum p_i \log_2 (p_i) \quad (1)$$

Keterangan :

S : Himpunan kasus;

N : Jumlah partisi S;

Pi : proporsi dari Si terhadap S.

Setelah mencari nilai entropi untuk suatu atribut selanjutnya mencari nilai gain untuk menentukan atribut mana yang digunakan buat menjadi node untuk pohon keputusan. Berikut rumus untuk gain :

$$(S, A) = Entropi (S) - \sum |S_i| |S| n_i = 1 X Entropi (S_i) \tag{2}$$

Keterangan

S : himpunan data;

A : atribut;

n : jumlah partisi atribut A;

|Si| : jumlah kasus pada partisi ke-I;

|S| : jumlah kasus dalam S.

Setelah semua atribut selesai dihitung menggunakan rumus diatas, maka nilai gain yang tertinggi yang akan menjadi node (akar) untuk pohon keputusan

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Algoritma C4.5

Perhitungan Algoritma C4.5 dimulai dengan pemilihan atribut akar yang terlebih dahulu mencari jumlah keseluruhan kasus, jumlah kasus ialah Tinggi dan Rendah. Dalam menghitung entrophy dari semua kasus yang dibagi berdasarkan Tahun 2019 sampai Tahun 2020. Setelah itu dilakukan perhitungan gain untuk masing – masing atribut.

a) Menghitung *entropy total* :

$$Entropy [Total] = \left(-\frac{15}{24} x \log_2 \left(\frac{15}{24} \right) \right) + \left(-\frac{9}{24} x \log_2 \left(\frac{9}{24} \right) \right)$$

$$Entropy [Total] = 0,954434003$$

b) Menghitung *entropy gain jumlah* :

$$Entropy [Jumlah - Banyak] = \left(-\frac{11}{11} x \log_2 \left(\frac{11}{11} \right) \right) + \left(-\frac{0}{11} x \log_2 \left(\frac{0}{11} \right) \right)$$

$$= 0$$

$$Entropy [Jumlah - Sedikit] = \left(-\frac{5}{13} x \log_2 \left(\frac{5}{13} \right) \right) + \left(-\frac{8}{13} x \log_2 \left(\frac{8}{13} \right) \right)$$

$$= 0,961236605$$

$$= 0,9544340 \left(\left(\frac{11}{24} x 0 \right) + \left(\frac{13}{24} x 0,961236605 \right) \right)$$

$$= 0,433764175$$

Berikut ini adalah table hasil perhitungan nilai entropy dan gain yang diuraikan pada tabel 2.

Tabel 2. Perhitungan Node 1

Node 1	Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Total	24	15	9	0,954434003	
Jumlah					0,433764175
	Banyak	11	11	0	
	Sedikit	13	5	0,961236605	
Harga/Kg					0,232328056
	Mahal	13	5	0,961236605	
	Murah	11	10	0,439496987	
Penjualan					0,02762912
	Produsen	11	8	0,845350937	
	Pedagang	13	7	0,995727452	
	Perantara				
Kemasan					0,112607355
	Daun	14	11	0,749595257	
	Plastik	10	4	0,970950594	

Dari hasil yang diperoleh pada tabel 2, nilai atribut yang tertinggi adalah Jumlah dengan gain sebesar 0,433764175. Maka atribut Jumlah di pilih sebagai node akar. Kelas atribut banyak adalah kosong sehingga tidak perlu perhitungan pada atribut tersebut. Untuk atribut banyak belum memperoleh hasil antara keputusan tinggi dan rendah.

Tabel 3. Perhitungan *Node 1.1*

Node 1.1		Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Jumlah-Sedikit		13	5	8	0,961236605	
Harga/Kg						0,532644292
	Mahal	9	2	7	0,764204507	
	Murah	4	3	1	0,811278124	
Penjualan						0,552288817
	Produsen	6	4	2	0,918295834	
	Pedagang Perantara	7	1	6	0,591672779	
Kemasan						0,452690504
	Daun	6	3	3	1	
	Plastik	7	2	5	0,863120569	

Dari hasil untuk perhitungan diatas pada tabel 3 atribut yang memiliki gain tertinggi adalah penjualan **0,552288817**. Dimana atribut dari penjualan yang terdiri dari produsen dan pedagang perantara yang masing-masing belum ditemukan hasilnya. Adapun hasil dari perhitungan selanjutnya dari algoritma C4.5 dapat dilihat dari table berikut ini.

Tabel 4. Perhitungan *Node 1.1.1*

Node 1.1.1		Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Jumlah-Sedikit, Penjualan-Produsen		6	4	2	0,918295834	
Harga/Kg						0,787767336
	Mahal	4	2	2	1	
	Murah	2	2	0	0	
Kemasan						0,724860044
	Daun	3	2	1	0,918295834	
	Plastik	3	2	1	0,918295834	

Dari hasil untuk perhitungan diatas pada tabel 4, atribut yang memiliki gain tertinggi adalah harga. Dimana atribut dari harga yang terdiri dari kelas didapatkan hasil tinggi sedangkan mahal belum diketahui. Maka diperlukan perhitungan lebih lanjut. Adapun hasil dari perhitungan selanjutnya dari algoritma C4.5 dapat dilihat dari table berikut ini

Tabel 5. Perhitungan *Node 1.1.1.1*

Node 1.1.1.1		Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Jlh-Sedikit, P-Produsen, Harga-Mahal		4	2	2	1	
Kemasan						0,787767336
	Daun	2	1	1	1	
	Plastik	2	1	1	1	

Tabel 6. Perhitungan *Node 1.1.2*

Node 1.1.2		Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Jlh-Sedikit, P-Pedagang Perantara		7	1	6	0,591672779	
Harga/Kg						0,87110067
	Mahal	5	0	5	0	
	Murah	2	1	1	1	
Kemasan						0,839647024
	Daun	3	1	2	0,918295834	
	Plastik	4	0	4	0	

Dari hasil untuk perhitungan diatas pada table 6 atribut yang memiliki gain tertinggi adalah harga. Dimana atribut dari harga yang terdiri dari kelas didapatkan hasil tinggi sedangkan mahal belum diketahui. Maka diperlukan perhitungan lebih lanjut.

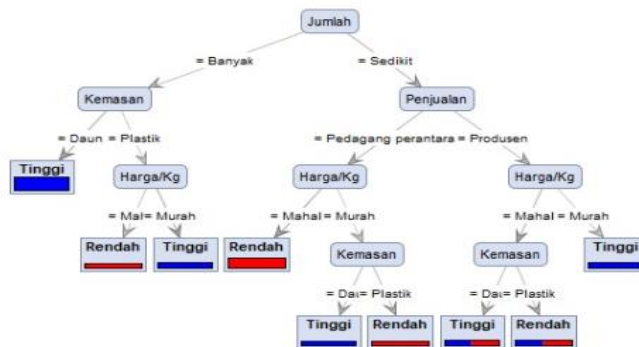
Tabel 7. Perhitungan *Node 1.1.2.1*

Node 1.2.1.1		Jml Kasus	Tinggi	Rendah	Entropy	Information Gain
Jlh-Sedikit, P-Pedagang Perantara, H-Murah		2	1	1	1	
Kemasan						0,954434003
	Daun	1	1	0	0	
	Plastik	1	0	1	0	

Dari hasil untuk perhitungan diatas pada tabel 7 hasil perhitungan **Node 1.1.2.1** diperoleh pohon atribut dengan hasil Daun = Tinggi dan Plastik = Rendah

3.2 Hasil Percobaan

Hasil pengolahan data dengan model pohon keputusan sesuai dengan *software rapidminer*, dapat dilihat pada gambar dibawah ini :



Gambar 1. Hasil *decision tree*

Selanjutnya dapat pada pola pohon keputusan akhir setelah melakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5 sebagai berikut :

```

Tree
Jumlah = Banyak
| Kemasan = Daun: Tinggi {Tinggi=8, Rendah=0}
| Kemasan = Plastik
| | Harga/Kg = Mahal: Rendah {Tinggi=0, Rendah=1}
| | Harga/Kg = Murah: Tinggi {Tinggi=2, Rendah=0}
Jumlah = Sedikit
| Penjualan = Pedagang perantara
| | Harga/Kg = Mahal: Rendah {Tinggi=0, Rendah=5}
| | Harga/Kg = Murah
| | | Kemasan = Daun: Tinggi {Tinggi=1, Rendah=0}
| | | Kemasan = Plastik: Rendah {Tinggi=0, Rendah=1}
| Penjualan = Produsen
| | Harga/Kg = Mahal
| | | Kemasan = Daun: Tinggi {Tinggi=1, Rendah=1}
| | | Kemasan = Plastik: Rendah {Tinggi=1, Rendah=1}
| | Harga/Kg = Murah: Tinggi {Tinggi=2, Rendah=0}
    
```

Gambar 2. Deskripsi *decision tree*

Dari pohon keputusan pada gambar 2 diperoleh *rules* sebagai berikut:

Tabel 8. *Rules* yang dihasilkan

No	Rules	Keputusan
1	Jika C1 = Banyak dan C4 =Daun	Tinggi
2	Jika C1 = Banyak dan C4 = Plastik dan C2=Mahal	Rendah
3	Jika C1 = Banyak dan C4 = Plastik dan C2=Murah	Tinggi
4	Jika C1=Sedikit dan C3=Pedagang Perantara dan C2=Mahal	Rendah
5	Jika C1=Sedikit dan C3=Pedagang Perantara dan C2=Murah dan C4=Daun	Tinggi
6	Jika C1=Sedikit dan C3=Pedagang Perantara dan C2=Murah dan C4=Plastik	Rendah
7	Jika C1=Sedikit dan C3=Produsen dan C2=Mahal dan C4=Daun	Tinggi
8	Jika C1=Sedikit dan C3=Produsen dan C2=Mahal dan C4=Plastik	Rendah
9	Jika C1=Sedikit dan C3=Produsen dan C2=Murah	Tinggi

Berikut ini tingkat akurasi yang didapatkan :

accuracy: 70.00% +/- 33.99% (mikro: 70.83%)			
	true Tinggi	true Rendah	class precision
pred. Tinggi	12	4	75.00%
pred. Rendah	3	5	62.50%
class recall	80.00%	55.56%	

Gambar 3. Nilai Akurasi Algoritma C4.5

```

PerformanceVector
PerformanceVector:
accuracy: 70.00% +/- 33.99% (mikro: 70.83%)
ConfusionMatrix:
True:  Tinggi Rendah
Tinggi: 12  4
Rendah:  3  5
classification_error: 30.00% +/- 33.99% (mikro: 29.17%)
ConfusionMatrix:
True:  Tinggi Rendah
Tinggi: 12  4
Rendah:  3  5

```

Gambar 4. Nilai *Performance Vector* Algoritma C4.5

Berdasarkan perhitungan pengolahan data menggunakan *software RapidMiner* mendapatkan nilai akurasi sistem sebesar 70.00%. Dimana model yang terbentuk sudah diuji tingkat akurasi dengan menggunakan data yang diuji berasal dari data training dengan *split validation* pada aplikasi *rapidminer 5.3*.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan pada Penerapan *Data Mining* dengan Algoritma C4.5 dalam Memprediksi Penjualan Tempe pada AC Mandiri dapat disimpulkan sebagai berikut. Permasalahan menentukan faktor dominan tinggi dan rendahnya hasil penjualan tempe dapat diselesaikan menggunakan teknik *data mining* dengan Algoritma C4.5. Menghasilkan 9 *rules* dengan tingkat akurasi mencapai 70%. Hasil perhitungan diperoleh pada faktor yang paling dominan adalah jumlah (C1) dengan nilai *gain* sebesar **0,433764175**.

REFERENCES

- [1] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota Sheih," *Data Min. Penjualan Prod. Dengan Metod. Apriori Pada Indomaret Galang Kota*, vol. 1, 2018.
- [2] F. A. Pambudi, A. P. Windarto, M. Fauzan, J. T. Hardinata, and R. Winanjaya, "Analisis Klasifikasi C4 . 5 Pada Pola Pembayaran Sepeda Motor Adira Cabang Pematangsiantar," vol. 1, no. 4, pp. 160–169, 2021.
- [3] P. Santoso and R. Setiawan, "Penerapan Metode Klasifikasi Decision Tree dan Algoritma C4.5 dalam Memprediksi Kriteria Nasabah Kredit Mega Auto Finance," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 200, 2020.
- [4] F. Rahman, H. Zulfia Zahro', and F. X. Ari Wibisono, "Penerapan Algoritma C4.5 Dalam Memprediksi Asal Calon Mahasiswa Berbasis Website (Studi Kasus : Fakultas Hukum Universitas Mataram)," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 4, no. 2, pp. 161–169, 2020.
- [5] D. A. R. Saragih, M. Safii, and D. Suhendro, "Penerapan Data Mining Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi di Program Studi Sistem Informasi," *J. Inf. Syst. Res.*, vol. 2, no. 2, pp. 173–177, 2021.
- [6] R. A. Saputra, "Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang," *Semin. Nas. Inov. dan Tren*, no. April, pp. 1–8, 2014.
- [7] T. Imandasari and A. P. Windarto, "Sistem Pendukung Keputusan dalam Merekomendasikan Unit Terbaik di PDAM Tirta Lihou Menggunakan Metode Promethee," *J. Teknol. dan Sist. Komput.*, vol. 5, no. 4, p. 159, 2017.
- [8] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4 . 5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Jatiji*, vol. 3, no. 2, pp. 95–103, 2017.
- [9] M. A. Puspa, "Implementasi Data Mining Klasifikasi Algoritma C4.5 Dalam Perekrutan Perangkat Desa," *Build. Informatics, Technol. Sci.*, vol. 1, no. 2, pp. 92–97, 2019.
- [10] W. D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.