

Data Mining Algoritma C4.5 menentukan Kelayakan Penerima Bantuan Covid 19

Angga Riadi¹, Sundari Retno Andani², Retno Rafiqi Dewi³

^{1,2,3}STIKOM Tunas Bangsa, Jl. Jend. Sudirman Blok A No.1-3, Pematangsiantar, Indonesia
anggariadi07@gmail.com

Abstrak. Dampak dari Pandemi COVID 19 dapat dilihat dari terpuruknya ekonomi masyarakat, khususnya pada masyarakat nagori Sitalasai tempat dilakukannya penelitian ini. Data dari Dinas Kelurahan menjelaskan bahwa hampir seluruh masyarakat merasakan sulitnya menjalankan roda perekonomian. Dalam mengatasi permasalahan ini pemerintah membuat program pemberian bantuan khusus bagi masyarakat yang layak. Untuk menentukan kelayakan penerimaan bantuan Covid 19 ini dibutuhkan model atau aturan agar tepat pada sasaran. Saat ini pihak kelurahan hanya memiliki data yang bersumber dari Dinas Sosial dan tidak semua masyarakat masuk kedalam daftar tersebut sehingga banyak masyarakat yang seharusnya layak menerima bantuan tidak masuk kedalam daftar penerima bantuan. Data mining dengan algoritma C4.5 merupakan salah satu metode didalam ilmu komputer untuk memecahkan permasalahan klasifikasi. Dengan algoritma C4.5 variabel untuk mengukur faktor-faktor kelayakan penerima bantuan ini dikorelasikan untuk menghasilkan aturan. Hasil yang akan dicapai pada penelitian ini adalah aturan pohon keputusan kelayakan penerima bantuan Covid 19. Dengan metode pemecahan permasalahan ini menggunakan data mining algoritma C4.5 dapat mengklasifikasikan kelayakan penerima bantuan Covid19 sehingga data yang diolah valid dan tepat sasaran.

Kata Kunci : Bantuan Covid19, Datamining, AlgoritmaC4.5, Klasifikasi, Pohon Keputusan.

Abstract. The impact of the COVID-19 Pandemic can be seen from the decline in the community's economy, especially in the Sitalasai nagori community where this research was conducted. Data from the Kelurahan Office explains that almost all people find it difficult to run the economy. In overcoming this problem, the government has created a program to provide special assistance to the people who deserve it. To determine the feasibility of receiving Covid 19 assistance, a model or rule is needed so that it is right on target. Currently, the kelurahan only has data sourced from the Social Service and not all communities are included in the list, so many people who should be eligible to receive assistance are not included in the list of beneficiaries. Data mining with the C4.5 algorithm is one of the methods in computer science to solve classification problems. With the C4.5 algorithm, the variables to measure the eligibility factors of these beneficiaries are correlated to produce rules. The results to be achieved in this study are the decision tree rules for the eligibility of recipients of Covid 19 assistance. With this problem solving method using data mining algorithm C4.5 can classify the eligibility of recipients of Covid 19 assistance so that the processed data is valid and on target.

Keyword : Covid19 Assistance, Datamining, AlgorithmC4.5, Classification, Decision Tree

PENDAHULUAN

Kelayakan penerima bantuan covid 19 merupakan salah satu aspek penting dalam membantu sistem pendukung keputusan bagi pangulu sitalasari dalam penetapan kesejahteraan bagi masyarakat. Penentuan kelayakan penerima bantuan perlu dilakukan dengan objektif agar bantuan yang diberikan tepat sasaran dan memenuhi azas keadilan bagi semua masyarakat[1]. Penetapan kelayakan penerima bantuan covid 19 dilakukan melalui proses meliputi jumlah penghasilan, status tempat tinggal, umur, dan jumlah tanggungan. Selain untuk menangani penyebaran covid 19 di lingkungan nagori Sitalasari, pemerintah setempat juga melaksanakan program pemerintah yaitu pemberian bantuan covid 19 kepada warga yang terdampak secara ekonomi di lingkungan nagori Sitalasari. Dalam menentukan penerima bantuan tersebut pihak kelurahan menggunakan data yang bersumber dari dinas sosial. Permasalahan dalam penentuan penerima bantuan tersebut adalah bahwa masih banyak warga dengan kondisi yang seharusnya layak tidak masuk kedalam daftar penerima bantuan. Hal ini tentu saja akan menjadi dilema bagi Pihak Kelurahan sebagai penyelenggara penyalur bantuan pemerintah. Dengan

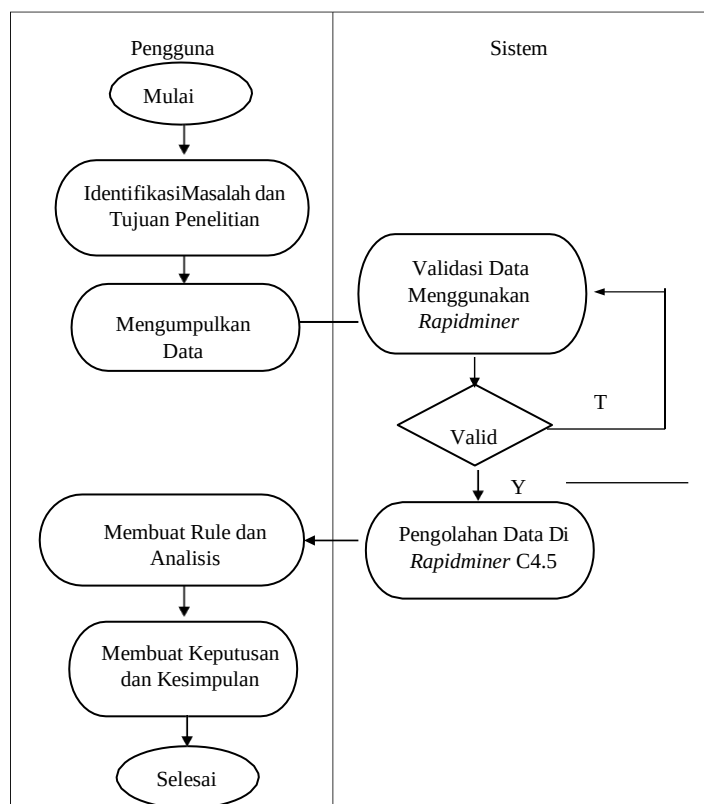


permasalahan tersebut dibuatlah penelitian ini dengan menganalisa data penduduk dan syarat penerima bantuan. Data mining merupakan salah satu keilmuan bidang komputer yang dapat membantu proses pembuat keputusan. Dengan algoritma C4.5 dapat mengatasi permasalahan klasifikasi[2],[3][4]. Algoritma C4.5 atau disebut juga sebagai algoritma decision tree merupakan metode klasifikasi dan prediksi. Algoritma ini termasuk metode untuk menemukan pola dengan memilah-milah sejumlah data yang besar menggunakan teknologi pengenalan pola. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya[5],[6].

Penerapan datamining dengan algoritma C4.5 pada permasalahan ini dapat membuat sistem aturan pohon keputusan sehingga dalam permasalahan kelayakan penerima bantuan Covid19 ini sesuai dengan aturan yang berlaku dan tepat sasaran[7].

METODOLOGI PENELITIAN

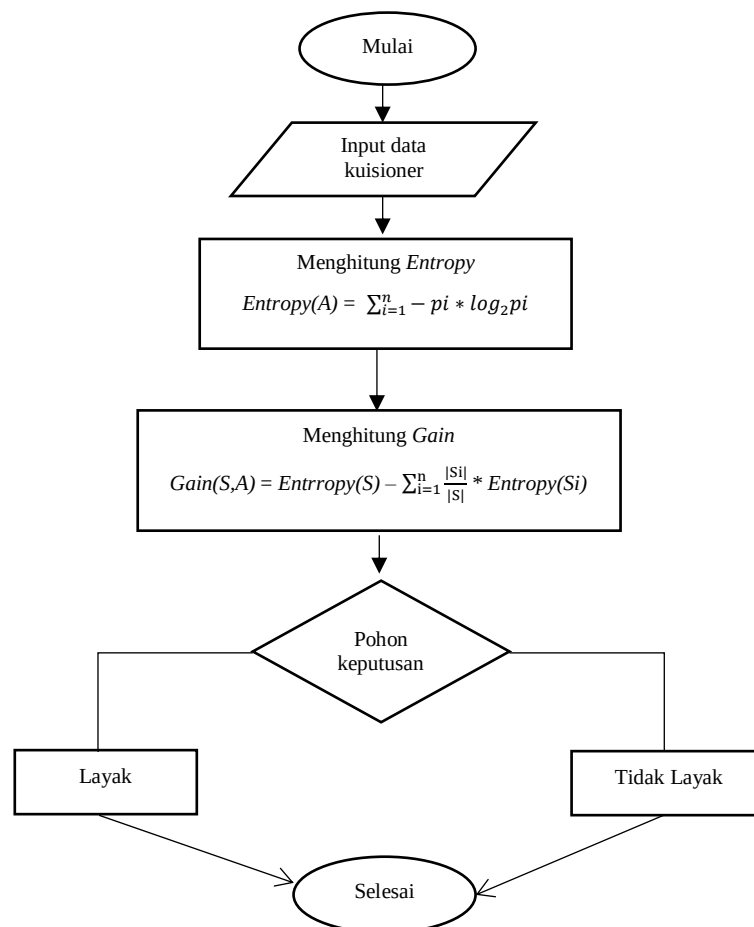
Penelitian ini dilakukan dengan observasi pada kantor pengulu sitalasai dengan objek data penduduk. Berdasarkan data tersebut maka dilakukan analisa masalah yang terkait dengan kelayakan penerima bantuan Covid19. Selanjutnya dilakukan studi literatur dengan merujuk pada beberapa penelitian atau jurnal yang berhubungan dengan pemecahan masalah menggunakan teknik datamining algoritma C4.5. Pada tahap analisa data selain data penduduk juga dibuat kuesioner kepada penduduk yang sesuai kriteria tetapi tidak masuk kedalam daftar penerima bantuan dengan atribut jumlah penghasilan, status tempat tinggal, usia dan jumlah tanggungan. Adapun aktifitas kerja pada penelitian ini adalah sebagai berikut:



Gambar 1. Aktifitas Penelitian

Agar didalam penelitian ini menghasilkan rule atau pohon keputusan yang sesuai dengan kelayakan penerimaan bantuan Covid19 ini maka dibuatlah pemodelan dengan implementasi

algoritma C4.5 berikut:



Gambar 2. Pemodelan Algoritma C4.5

Metode yang digunakan pada penelitian ini adalah algoritma C4.5[8] [9]. Dalam pemodelan algoritma C4.5 dilakukan pada sampel dataset, kemudian menghitung *entropy* (S) dari keseluruhan atribut, setelah *entropy* (S) ditemukan kemudian menghitung *Gain* tertinggi dari seluruh atribut, atribut yang memiliki *Gain* tertinggi yang akan digunakan sebagai akar/node. Selanjutnya buat cabang untuk tiap nilai, bagi kasus dalam cabang, ulangi perhitungan *Gain* sampai semua data masuk kedalam kelas yang sama[10],[5]. Atribut yang telah dipilih tidak lagi dibutuhkan dalam perhitungan, proses pembentukan pohon keputusan terhenti jika sudah tidak ada lagi atribut yang dipartisi dan semua tuple dalam *node* N telah memiliki kelas yang sama[11]. Dalam kasus dataset pada penelitian ini terdiri dari 2 kelas keterangan yaitu Layak dan Tidak Layak. Berikut adalah digram alur kerja pemodelan Algoritma C4.5.

HASIL DAN PEMBAHASAN

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan layak dan jumlah kasus keputusan tidak layak. Menghitung *entropy* dari semua kasus yang dibagi berdasarkan jumlah penghasilan (C1), status tempat tinggal (C2), umur (C3), dan jumlah tanggungan (C4). Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Untuk perhitungan Algoritma C4.5 dapat diuraikan sebagai berikut :



- A. Menghitung jumlah kasus, jumlah kasus untuk keputusan layak, dan jumlah kasus untuk keputusan tidak layak.

Menghitung Entropy total :

$$Entropy (total) = - (73/100) * \text{LOG}_2(73/100) - (27/100) * \text{LOG}_2(27/100)$$

$$Entropy (total) = 0,841464636$$

- B. Menghitung *Entropy* dan *gain* untuk masing-masing atribut data. Berikut ini adalah perhitungan nilai *Entropy* dan *Gain*. Menghitung *Entropy* dan *Gain* jumlah penghasilan (C1) :

Entropy [C1-rendah]

$$= -(0/43) * \text{LOG}_2(0/43) - (43/43) * \text{LOG}_2(43/43) = 0$$

Entropy [C1-cukup]

$$= -(0/26) * \text{LOG}_2(0/26) - (26/26) * \text{LOG}_2(26/26) = 0$$

Entropy [C1-tinggi]

$$= -(27/31) * \text{LOG}_2(27/31) - (4/31) * \text{LOG}_2(4/31) = 0,554778163$$

Gain [total-C1]

$$= 0,841464636 - ((43/100) * 0 + (26/100) * 0 + (31/100) * 0,554778163)$$

$$= 0,669483406$$

Menghitung *Entropy* dan *Gain* status tempat tinggal (C2):

Entropy [C2-kontrak]

$$= -(18/62) * \text{LOG}_2(18/62) - (44/62) * \text{LOG}_2(44/62) = 0,869137581$$

Entropy [C2-rumah sendiri]

$$= -(9/38) * \text{LOG}_2(9/38) - (29/38) * \text{LOG}_2(29/38) = 0,789749254$$

Gain [total-C2]

$$= 0,841464636 - ((62/100) * 0,869137581 + (38/100) * 0,789749254)$$

$$= 0,00249462$$

Menghitung *Entropy* dan *Gain* umur (C3):

Entropy [C3-rendah]

$$= -(4/21) * \text{LOG}_2(4/21) - (17/21) * \text{LOG}_2(17/21) = 0,702466551$$

Entropy [C3-cukup]

$$= -(23/71) * \text{LOG}_2(23/71) - (48/71) * \text{LOG}_2(48/71) = 0,908618598$$

Entropy [C3-tinggi]

$$= -(0/8) * \text{LOG}_2(0/8) - (8/8) * \text{LOG}_2(8/8) = 0$$

Gain [total-C3]

$$= 0,841464636 - ((21/100) * 0,702466551 + (71/100) * 0,908618598 + (8/100) * 0)$$

$$= 0,048827456$$

Menghitung *Entropy* dan *Gain* jumlah tanggungan (C4):

Entropy [C4-rendah]

$$= -(17/17) * \text{LOG}_2(17/17) - (0/17) * \text{LOG}_2(0/17) = 0$$

Entropy [C4-cukup]

$$= -(10/63) * \text{LOG}_2(10/63) - (53/63) * \text{LOG}_2(53/63) = 0,631263018$$

Entropy [C4-tinggi]

$$= -(0/20) * \text{LOG}_2(0/20) - (20/20) * \text{LOG}_2(20/20) = 0$$

Gain [total-C3]

$$= 0,841464636 - ((17/100) * 0 + (63/100) * 0,631263018 + (20/100) * 0)$$

$$= 0,0443768935$$

Berikut ini hasil perhitungan nilai Entropy dan Gain yang dapat dilihat pada tabel 4.1.

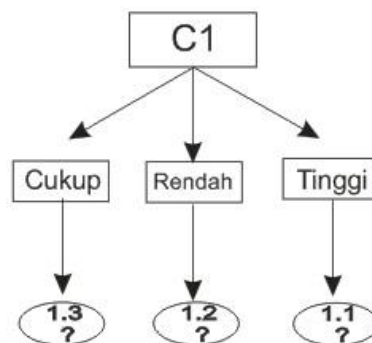
Tabel 1. Perhitungan Node 1

Total		Jumlah	Layak	Tidak Layak	Entropy	Gain
-------	--	--------	-------	-------------	---------	------



		100	73	27	0.841464636	
C1						0.669483406
	Rendah	43	43	0	0	
	Cukup	26	26	0	0	
	Tinggi	31	4	27	0.554778163	
C2						0.00249462
	Kontrak	62	44	18	0.869137581	
	Rumah sendiri	38	29	9	0.789749254	
C3						0.048827456
	Rendah	21	17	4	0.702466551	
	Cukup	71	48	23	0.908618598	
	Tinggi	8	8	0	0	
C4						0.443768935
	Rendah	17	0	17	0	
	Cukup	63	53	10	0.631263018	
	Tinggi	20	20	0	0	

Dari hasil perhitungan pada tabel 1 diperoleh atribut yang menjadi *node* (akar) adalah C1 memiliki gain tertinggi yaitu 0,669483406, dimana terdiri dari 3 sub atribut yaitu rendah, cukup, dan tinggi. Berdasarkan nilai *entropy* dari ketiga sub atribut diatas, sub atribut rendah dan cukup yang memperoleh keputusan layak. Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar selanjutnya, dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 2. Pohon Keputusan 1

Untuk hasil perhitungan selanjutnya dapat ditunjukkan pada tabel berikut ini:

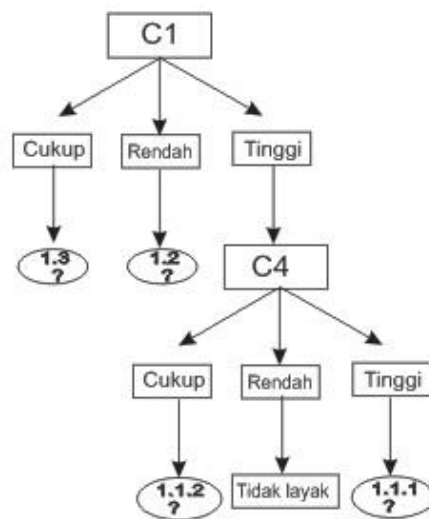
Table 2. Perhitungan Node 1.1

total		Jumlah	Layak	Tidak Layak	Entropy	Gain
C1=tinggi		31	4	27	0.554778163	
C2						0.009477045
	Kontrak	20	2	18	0.468995594	
	Rumah sendiri	11	2	9	0.684038436	
C3						0.12204848
	Rendah	4	0	4	0	
	Cukup	26	3	23	0.51594693	
	Tinggi	1	1	0	0	
C4						0.303156581



total C1=tinggi		Jumlah	Layak	Tidak Layak	Entropy	Gain
		31	4	27	0.554778163	
	Rendah	17	0	17	0	
	Cukup	12	2	10	0.650022422	
	Tinggi	2	2	0	0	

Dari hasil perhitungan diatas pada tabel 2 atribut yang menjadi *Node* cabang dari atribut C1 = tinggi adalah C4. Dimana atribut dari C4 yang terdiri dari Rendah, cukup, dan tinggi. Dimana rendah sudah memperoleh keputusan tidak layak. Sedangkan untuk sub atribut cukup dan tinggi belum memperoleh keputusan maka akan dilakukan perhitungan kembali. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



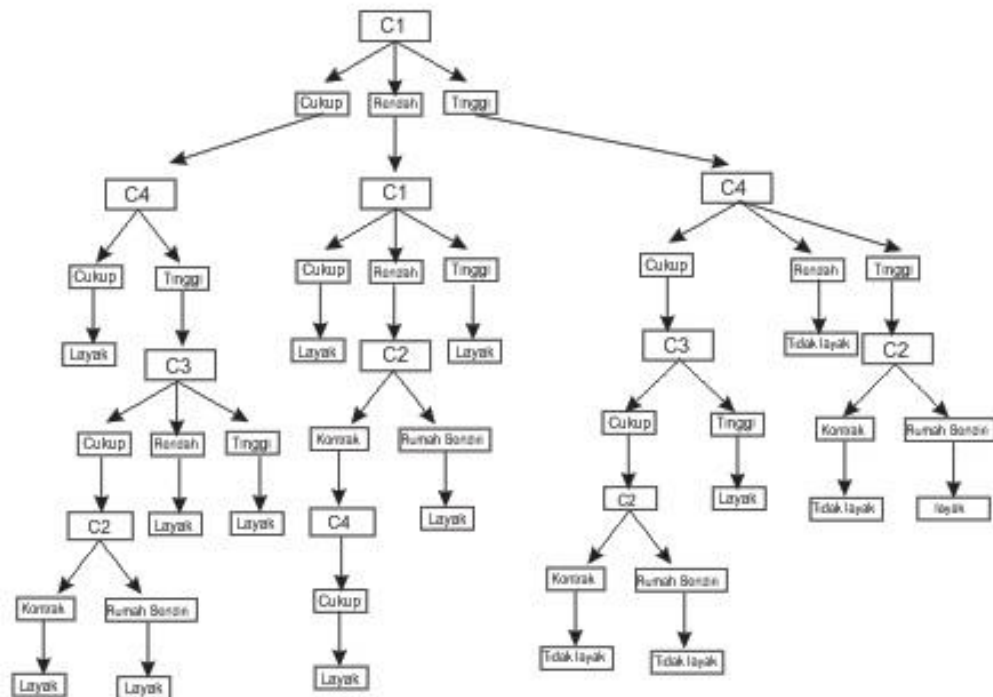
Gambar 3. Pohon Keputusan 2

Hasil akhir dari perhitungan variabel-variabel terkait adalah perhitungan pada *node* cabang dari atribut C1=Cukup, C4=tinggi, dan C3=cukup yang ditunjukkan pada table 3.

Table 3. Perhitungan Node 1.3.1.1.

total C1=Cukup C4=tinggi C3=Cukup		Jumlah	Layak	Tidak Layak	Entropy	Gain
		7	7	0	0	
C2						0
	Kontrak	1	1	0	0	
	Rumah sendiri	6	6	0	0	

Dari hasil perhitungan pada tabel 3, Dimana C2 kontrak dan rumah sendiri memperoleh keputusan layak. Maka dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :



Gambar 4. Pohon Keputusan 11

KESIMPULAN

Dari hasil penelitian yang dilakukan, maka kesimpulan yang dapat diambil berdasarkan pembahasan Untuk Menentukan Penerima Bantuan Covid 19 Menggunakan Algoritma C4.5 bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan kelayakan penerima bantuan Covid 19 adalah jumlah penghasilan. Hal ini ditunjukkan dengan atribut jumlah penghasilan yang menempati sebagai simpul akar (*rootnode*) dalam diagram pohon keputusan, dan atribut Jumlah tanggungan sebagai factor kedua dalam menentukan kelayakan penerima bantuan Covid 19. Hal ini dapat dilihat dalam diagram pohon keputusan bahwa atribut Jumlah tanggungan terletak dalam simpul cabang di bawah jumlah penghasilan dan berdasarkan hasil pengujian diperoleh model aturan pohon keputusan yang dapat digunakan untuk menentukan kelayakan penerima bantuan sosial khususnya bantuan Covid19.

DAFTAR PUSTAKA

- [1] TNP2K, "PROGRAM BANTUAN PEMERINTAH Menuju Bantuan Sosial Terintegrasi," p. 210, 2018, [Online]. Available: website TNP2K (www.tnp2k.go.id).
- [2] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 244–250, 2018, doi: 10.33096/ilkom.v10i2.300.244-250.
- [3] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.
- [4] M. F. Arifin and D. Fitrianah, "Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *IncomTech*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.
- [5] D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System On Students Using C4.5 Algorithm," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol.



- 19, no. 2, pp. 358–365, 2020, doi: 10.30812/matrik.v19i2.685.
- [6] S. Wahyuni, “Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree,” *J. Phys. Conf. Ser.*, vol. 970, no. 1, 2018, doi: 10.1088/1742-6596/970/1/012030.
- [7] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [8] E. Buulolo, K. Medan, and S. Utara, “C4 . 5 A lgorithm to Predict the Impact of the Earthquake,” vol. 6, no. 02, pp. 10–15, 2017.
- [9] A. A. Aprilia Lestari, “Increasing Accuracy of C4 . 5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease,” *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 32–38, 2020.
- [10] H. Amalia, ; Yunita, ; Ari Puspitasari, ; Ade, and F. Lestari, “Student Performance Analysis Using C4.5 Algorithm To Optimize Selection,” *J. PILAR Nusa Mandiri*, vol. 16, no. 2, pp. 149–154, 2020.
- [11] A. Andriani, “Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout,” *Semin. Nas. Mat. 2012*, pp. 139–147, 2012.

