

Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 Kota Pagar Alam

Rega Satya Putra^(✉)

Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
regasatyaputra66@gmail.com

Erwin Dwika Putra², M. Husni Rifqo³, Harry Witriyono⁴

Universitas Muhammadiyah Bengkulu, Bengkulu, Indonesia
erwindwikap@umb.ac.id², mhrifqo@umb.ac.id³,
harrywitriyono@umb.ac.id⁴

Abstrak—Pada awal tahun 2020 nyaris seluruh Negara didunia tidak terkecuali Indonesia menghadapi pandemi Covid-19 ataupun virus corona. Resiko penularan covid- 19 bisa dihindari dengan kedisiplinan masyarakat dengan mengurangi kegiatan ditengah keramaian jika tidak berkepentingan dan menerapkan protokol kesehatan. Kemajuan teknologi informasi memudahkan masyarakat memperoleh data dengan mudah yang cenderung berlebihan, kegiatan pengklasifikasian yang dilakukan manusia masih memiliki keterbatasan. Pada penelitian ini akan dibangun sistem klasifikasi untuk klasifikasi data covid-19 kota Pagar Alam menggunakan Algoritma C4.5 dengan metode CRIPS-DM dengan model *operators X-cross validation*. Hasil yang diperoleh dalam pengujian dengan menerapkan algoritma C4.5 pada data covid-19 kota Pagar Alam adalah Dari hasil pengujian dengan menerapkan algoritma C4.5 pada *rapid miner* diperoleh hasil *Accurasy* sebesar 86.67% dan hasil AUC (*Optimiscal*) diperoleh 0.989.

Abstract—All of countries in the world, including Indonesia faced the Covid-19 pandemic or the corona virus at the beginning of 2020. The risk of Covid-19 transmission can be avoided with public discipline by reducing activities in the middle of a crowd if there is no need and implementing health protocols. The development of information technology make it easier for people to easily obtain data that tends to be redundant and classification activities that carried out by humans still have limitations, especially the human ability to accommodate data to be classified. This study aimed to built the the classification of covid-19 data by using the C4.5 Algorithm in the city of Pagar Alam with the CRIPS-DM method with the X-cross validation operators model. The results obtained in testing by applying the C4.5 algorithm to the Covid-19 data for the city of Pagar Alam show that the accuracy results are 86.67% and the AUC (Optimise) results are 0.989.

Keywords—*Classification, Data Mining, and C4.5 Algorithm*

1. Pendahuluan

Pada awal tahun 2020 nyaris seluruh Negara di dunia tidak terkecuali Indonesia menghadapi pandemi Covid-19 ataupun virus corona. Bertepatan pada 11 Maret

2020 *World Health Organization* menetapkan covid- 19 ataupun virus corona sebagai pandemi [1]. Resiko penularan covid- 19 bisa dihindari dengan kedisiplinan masyarakat dengan mengurangi kegiatan ditengah keramaian jika tidak berkepentingan dan menerapkan protokol kesehatan seperti yang telah dianjurkan pemerintah. Kemajuan teknologi informasi memudahkan masyarakat memperoleh data dengan mudah yang cenderung berlebihan, kegiatan pengklasifikasian yang dilakukan manusia masih memiliki keterbatasan, terutama kemampuan manusia menampung data yang akan diklasifikasi banyaknya data yang masuk pada suatu lembaga atau organisasi setiap harinya yang semakin lama akan semakin banyak, tumpukan data akan terakumulasi sehingga pemanfaatnya tidak optimal. Banyaknya data tersebut membuat proses pengklasifikasian terhambat dan mengalami kekeliruan [2].

Penelitian sebelumnya oleh Turnip, S. Dkk [3]. Membahas tentang Analisis Pola Penyebaran Penyakit Dengan Menggunakan Algoritma C4.5 memiliki akurasi yang sangat baik dalam klasifikasi penyakit tuberkolosis. Maka dalam peneltian ini akan menerapkan Algoritma C4.5 pada kasus covid-19 di kota Pagar Alam, penambahan metode CRIPS-DM diharapkan agar dapat meningkatkan kinerja algoritma C4.5 dan menghasilkan informasi yang akurat.

2. Studi Literatur

2.1 Covid-19

Corona virus atu dengan istilah covid-19 adalah termasuk dalam kategori virus yang dapat menyebabkan penyakit pada manusia dan hewan. Berdasarkan banyak kasus yang terjadi pada manusia virus covid-19 biasanya menyebabkan penyakit infeksi saluran pernafasan seperti flu hingga penyakit serius seperti *middle east respiratory syndrome* (MERS) dan penyakit pernafasan akut *savere acute resviratory syndrome* (SARS). Penyebaran covid-19 bisa terjadi melalui tetesan kecil (*droplet*) dari hidung atau mulut disaat batuk bersin, *droplet* keluar bertebaran diudara terhirup atau jatuh menempel pada benda yang ada disekitarnya kemudian benda yang terkontaminiasi *droplet* tersebut tersentuh oleh orang yang ada disekitar. 11 Maret 2020 Badan Kesehatan Dunia atau *World Health Organization* menetapkan covid- 19 ataupun virus corona sebagai pandemi global [1]

2.3 Klasifikasi

Klasifikasi adalah teknik *data mining* yang mengelompokkan data kedalam kelas atau kelompok tertentu, proses penemuan jenis data baru yang bisa menentukan kelompok baru dan membedakan kelas atau kelompok data sehingga dapat dimanfaatkan untuk mengambil keputusan dan memprediksi kelas objek yang diteliti dan menentukan label kelas yang belum diketahui [2]

2.4 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk mengklasifikasi *datashet* dengan membuat pohon keputusan yang dapat dimanfaatkan untuk memprediksi suatu keputusan sesuai dengan aturan pengambilan keputusan. Dalam pohon keputusan merupakan diagram alir dan setiap internal node menghimpun atribut yang akan diuji, pada setiap cabang mempresentasikan kelas-kelas tertentu [3]. Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai gain maka kita perlu mencari nilai *entropy* terlebih dahulu, rumus yang digunakan berikut ini;

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

Sementara untuk menentukan nilai gain dapat dilihat pada persamaan berikut;

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

2.5 Rapid Miner

Rapid Miner adalah software atau perangkat lunak yang digunakan untuk mengolah data berdasarkan prinsip dan algoritma *data mining*. Dari *data set* Rapid miner mengekstrak pola-pola, menggabungkan metode stastika dengan kecerdasan buatan dan *data base* [4].

2.6 Penelitian Terkait

Penelitian yang dilakukan oleh Turnip, S. M dkk, dengan judul “Analisis Pola Penyebaran Penyakit Dengan Menggunakan Algoritma C4.5” algoritma C4.5 mampu mengelompokan penyakit yang dominan terjadi, sehingga dapat dijadikan pihak rumah sakit untuk melakukan sosialisasi dan pencegahan [3].

Pada penelitian yang dilakukan Septiani dengan judul “Komperasi Metode Klasifikasi *Data Mining Algoritma C4.5* dan *Naive Bayes* Untuk Memprediksi Penyakit Hepatitis” dari hasil penelitian kedua metode ini akurat dalam melakukan prediksi untuk penyakit hepatitis [5].

Penelitian sebelumnya oleh Bastian Dkk, dengan judul “Penerapan Algoritma K-Means *Clustering* Anaysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)” Algoritma K-means *clustering data mining* dapat digunakan dalam penanganan penyakit menular di kabupaten Majalengka dan bisa mengontrol distribusi obat-obatan kesetiap puskesmas dan meningkatkan pelayanan dengan mempersiapkan SDM dan perlengkapan medis bila diperlukan ke poskesmas yang memiliki pasien penyakit menular terbanyak [6].

3. Metode Penelitian

Dalam penelitian ini data yang digunakan adalah data jumlah penduduk kota pagar alam tahun 2020 yang diperoleh langsung pada Dinas Kependudukan dan Pencatatan Sipil kota Pagar Alam dan data kasus Covid-19 Kota Pagar Alam *update* 11 Januari 2021 dari Gugus Tugas Covid-19 melalui Dinas Kesehatan kota Pagar Alam. Metode penelitian ini menggunakan teknik *data mining* yaitu *Cross Industry Standard Process Data Mining* (CRISP-DM) [7]. Tahapan metode CRSP-DM dapat dilihat pada gambar berikut ini.

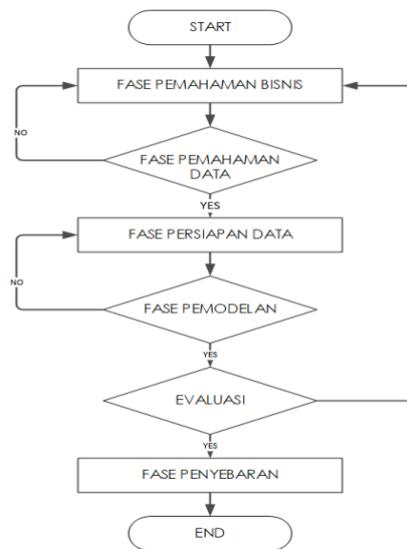


Fig. 1. Flowchart tahapan penelitian metode CRISP-DM

1. Fase pemahaman bisnis (*business understanding phase*)
Pada tahapan pemahaman bisnis ada beberapa hal yang dilakukan antara lain, menentukan tujuan bisnis, menilai situasi, dan menentukan tujuan data mining.
2. Fase pemahaman data (*data understanding phase*)
Pada tahapan pemahaman data, beberapa hal yang perlu dilakukan antara lain, mengumpulkan data awal, mendeskripsikan data, mengeksplorasi data, dan memverifikasi kualitas data.
3. Fase persiapan data (*data preparation phase*) Pada tahapan persiapan data, teknik persiapan data agar kualitas data diperoleh lebih baik dengan cara pembersihan data (*data cleaning*) dan integrasi data (*data integration*), Seleksi data (*data selection*) dan transformasi data (*data transformation*)
4. Pemodelan Pada penelitian yang menggunakan Algoritma C4.5 dengan *operator Cross Validation* dan dengan metode pemilihan *split* menggunakan *gain ratio*. Untuk melakukan pengukuran dalam akurasi penelitian ini akan menggunakan *software RapidMiner 8.9*.
5. Evaluasi (*evaluation*)

Pada tahap ini dilakukan evaluasi terhadap keefektifan dan keakuratan dari model yang terbentuk sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase pemahaman bisnis.

6. Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan dipresentasikan dalam bentuk laporan sehingga dapat diketahui atau dibaca oleh orang awam. Hasil dari penelitian ini berupa klasifikasi penyebaran covid-19 di kota Pagar Alam.

4. Hasil dan Pembahasan

Data yang digunakan dalam pengolahan data data ini adalah data yang sudah melalui proses pembersihan data, selanjutnya dalam tahap ini akan dilakukan penyelesaian masalah pada klasifikasi penyebaran covid-19 menggunakan algoritma C4.5 kota Pagar Alam dengan metode CRSIP-DM.

4.1. Perhitungan Algoritma C4.5

Pada perancangan klasifikasi dengan *data mining* menerapkan algoritma C4.5 pada penelitian ini menghasilkan pohon keputusan dengan proses pengubahan data yang awalnya berbentuk tabel untuk diolah menjadi pohon keputusan. Langkah awal dalam membuat pohon keputusan yaitu menghitung jumlah kasus keseluruhan Rawan dan Tidak Rawan Covid-19 berdasarkan atribut yang digunakan yaitu, Kecamatan, Jenis Kelamin, Usia, Kasus Positif Covid-19, Kasus Negatif Covid-19, Meninggal Dunia, dan Dampak Covid-19. Selanjutnya menghitung nilai *entropy*, kemudian lanjutkan dengan menghitung nilai *entropy* dan *gain* tertinggi pada setiap nilai-nilai dalam variabel. Pada penelitian ini menggunakan *gain ratio* makan pertama mencari nilai *gain ratio* sesuai dengan alur algoritma C4.5 yaitu menentukan *entropy* dan *gain* terlebih dahulu.

Langkah 1, menghitung *Entropy* total dari masing-masing nilai kasus covid-19 kota Pagar Alam berikut ini.

$$\begin{aligned} EntropyTota (Rawan^{23}, TidakRawan^7) &= (-23/30 * \log_2(23/30)) + (-7/30 * \log_2(7/30)) \\ &= 0,783776947 \end{aligned}$$

Langkah 2, Menghitung *Entropy* dari tiap-tiap atribut Kecamatan

$$\begin{aligned} EntropyPagarAlamSelatan (Rawan^6, TidakRawan^0) \\ &= (-6/6 * \log_2(6/6)) + (-0/6 * \log_2(0/6)) \\ &= 0 \end{aligned}$$

Ulangi perhitungan *entropy* seperti diatas pada setiap atribut Kecamatan, kemudian

Hitung *Entropy* total Kecamatan

$$\begin{aligned} \text{EntropyTotal Kecamatan} &= 0 + 0 + 0 + 6/30 * (0,650022422) + 0 \\ &= 0,130004484 \end{aligned}$$

Langkah ke 4, hitung *Split Information*, dengan rumus persamaan berikut.

$$\begin{aligned} \text{Split information (Total Kecamatan)} &= (-6/30 * \log_2(6/30) + (-6/30 * \log_2(6/30) + \\ &\quad (-6/30 * \log_2(6/30) + (-6/30 * \log_2(6/30) + \\ &\quad (-6/30 * \log_2(6/30) + \\ &= 2,321928095 \end{aligned}$$

Langkah ke 5, hitung *Gain Ratio*, dengan rumus persamaan berikut.

$$\begin{aligned} \text{Gain Ratio(Total, Kecamatan)} &= 0,653772463 / 2,321928095 \\ &= 0,281564474 \end{aligned}$$

Ulangi perhitungan seperti langkah-langkah diatas pada setiap atribut yang digunakan sehingga mendapatkan hasil *gain ratio* seperti pada table 1 dibawah ini;

Tabel 1. Hasil Perhitungan *Root/node 1*

Atribut	Nilai	Total	Rawan	Tidak Rawan	Entropy	Gain	SplitInfo	Gain Ratio
Kecamatan		30	23	7	0,783776947	0,130004484	2,321928094	0,21564474
	Pagar Alam Selatan	6	6	0	0			
	Pagar Alam Utara	6	6	0	0			
	Dempo Selatan	6	0	6	0			
	Dempo Tengah	6	5	1	0,6500224221			
	Dempo Utara	6	6	0	0			
Jenis Kel	L	15	11	4	0,836640742	0,004492529	1	0,004492529
	P	15	12	3	0,721928095			
Usia	Anak	10	7	3	0,881290899	0,008727918	1,584962501	0,005506703
	Muda	10	8	2	0,721928095			
	Tua	10	8	2	0,721928095			
Positif Covid-	0	11	11	7	0,945660305			
	1	1	1	0				
	2	5	5	0				
	3	3	3	0				

	6	1	1	0		0,437034836	2,89294807	0,151069022
	7	1	1	0				
	8	2	2	0				
	9	2	2	0				
	11	1	1	0				
	12	2	2	0				
	14	1	1	0				
Negatif Covid-19	1	1	1	0		0,437034836	2,89294807	0,151069022
	3	2	2	0				
	4	1	1	0				
	6	2	2	0				
	7	2	2	0				
	8	1	1	0				
	9	1	1	0				
	12	3	3	0				
	13	5	5	0				
	14	1	1	0				
Meninggal Dunia	15	11	4	7	0,945660305	0,087085372	0,972253928	0,089570605
	0	24	17	7	0,870864469			
	1	4	4	0				
	2	1	1	0				
	3	1	1	0				
Dampak Covid-19	Terdampak	6	6	0		0,437034836	1,517915957	0,287917677
	Sangat Terdampak	13	13	0				
	Tidak Terdampak	11	4	7	0,945660305			

Dari hasil perhitungang manual penentuan *root* atau akar pohon keputusan, yang memiliki nilai *gain ratio* tertinggi adalah variable Dampak Covid-19 dengan nilai *gain ratio* 0,287917677 dan dijadikan *root*. Dalam atribut Dampak Covid-19 memiliki nilai yaitu, “Terdampak”, “Sangat Terdampak,” dan “Tidak Terdampak” dapat dilihat dari hasil perhitungan bahwa untuk atribut Dampak Covid-19 dengan nilai “Terdampak” dan “Sangat Terdampak” masing- masing nilai seluruhnya memiliki 6 kategori rawan sehingga dapat disimpulkan kedua nilai tersebut menjadi daun atau *leaf*. Berikut adalah hasil pohon keputusan menggunakan hitung manual pada gambar 2 berikut:



Fig. 2. Pohon Keputusan perhitungan manual

4.2. Hasil Pengukuran Algoritma C4.5 Rapid Miner

Data yang telah diolah menjadi data *sheet* yang siap olah dilakukan pengujian pada aplikasi *Rapid Miner* dengan menerapkan Algoritma yang diusulkan yaitu Algoritma C4.5. *Operators* yang digunakan adalah *X-Cross Validation* dengan perhitungan *gain ratio*, alasan pengujian menggunakan model oprator ini adalah diharapkan hasil pengujian nantinya mendapatkan akurasi yang maksimal, berikut gambar 3 *X-Operators X-Cross Validation*.

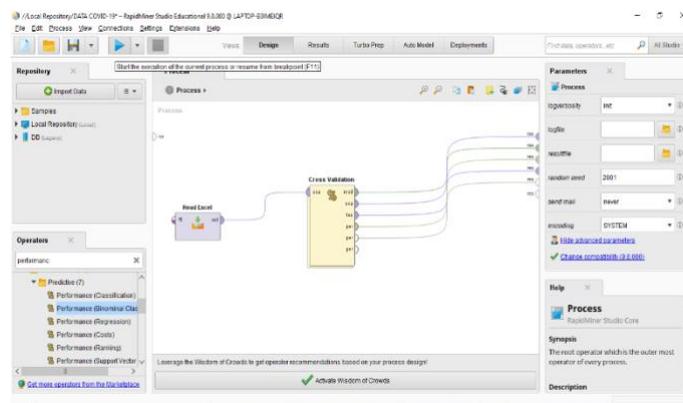


Fig. 3. Tampilan Read Excel dan Validation pada main process

Ketikkan *X-valuation* pada kolom pencarian di *Operators* dan *drop* ke *process* Pada kotak *Validation* klik dua kali sehingga muncul halaman *process* untuk *training* dan *testing* seperti gambar 4 berikut ini.

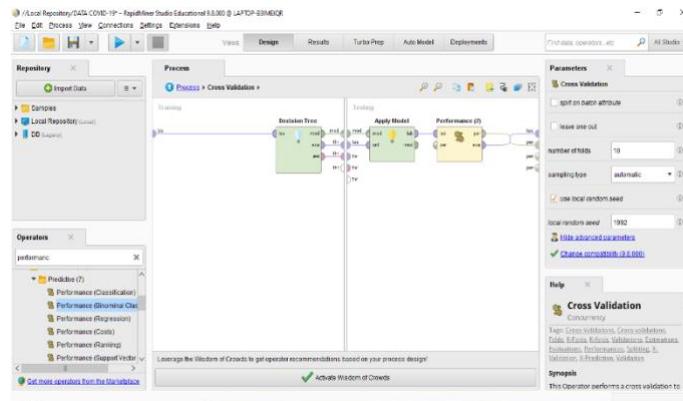


Fig. 4. Tampilan Proses Validation

Dari hasil pengujian dengan menerapkan algoritma C4.5 pada *rapid miner* diperoleh hasil *Accuracy* sebesar 86.67%. Dengan hasil *confusion matrix* dalam tabel 2 berikut ini.

Tabel 2. Hasil *Confusion Matrix*

	<i>True RAWAN</i>	<i>True TIDAK RAWAN</i>
Pred. RAWAN	20	1
Pred. TIDAK RAWAN	3	6

Hasil pengujian data yang dihasilkan oleh *Rapid Miner* dengan menggunakan model algoritma C4.5 adalah *accuracy*, *AUC (optimistic)*, *AUC (pessimistic)*, *precision*, dan *recall* pada gambar 5 berikut ini;

1. *Accuracy*

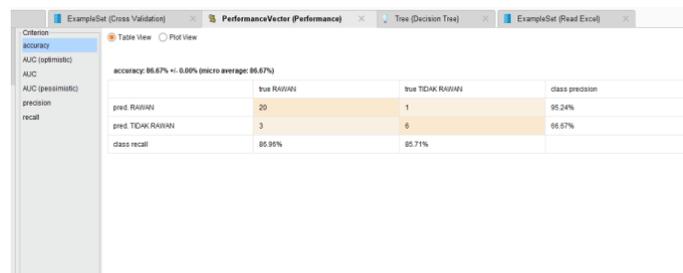


Fig. 5. Hasil *Accuracy*

Berdasarkan hasil pengukuran akurasi data yang mencapai 86.67%. Nilai

prediksi rawan dengan *true* rawan menjadi 20 prediksi dan *true* tidak rawan 1, dengan hasil nilai presisi mencapai 95.24%. Sedangkan untuk prediksi tidak rawan mencapai 3, dan untuk nilai *true* tidak rawan menjadi 6 dengan memperoleh hasil pencapaian sebesar 66.67%. untuk *recall* data keterangan yang paling tinggi ada pada *class recall true* tidak rawan sebesar 86.96%. Sedangkan untuk nilai *recall true* rawan sebesar 85.71%.

4 ROC/AUC (Area Under Cover)

Performance akurasion AUC (Gorunescu, 2010) dapat dikelompokkan menjadi lima kelompok yaitu:

- 0.90 – 1.00 = *Excellent Classification*
- 0.80 – 0.90 = *Good Clasification*
- 0.70 – 0.80 = *Fair Clasification*
- 0.60 – 0.70 = *Poor Clasificatio*
- 0,50 – 0.60 = *Failure*.

1. ROC/AUC (Optimiscal)



Fig. 6. ROC/AUC (Optimistic)

Hasil AUC (*Optimiscal*) diperoleh pada gambar di atas adalah 0.989 +/- 0.016 (*micro average: 0.989*) (*Positive class: TIDAK RAWAN*) dan termasuk kedalam tingkatan akurasi *Good Clasification*. Dari hasil pengujian dengan menggunakan *Rapid Miner* dengan menerapkan algoritma C4.5 diperoleh pohon keputusan seperti gambar 7 dibawah ini:

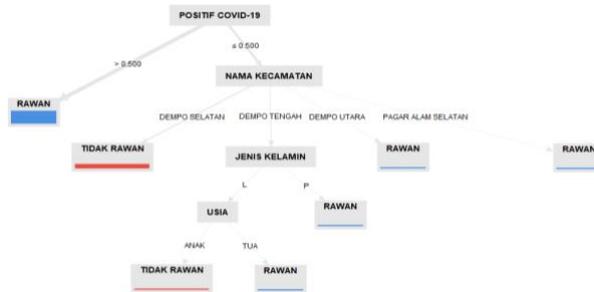


Fig. 7. Hasil Pohon Keputusan C4.5

Dari pohon keputusan diatas maka hasil pengujian dengan *rapidminer* memperoleh *rule* dihasilkan seperti gambar 8 sebagai berikut.

Tree

```

POSITIF COVID-19 > 0.500: RAWAN {RAWAN=19, TIDAK RAWAN=0}
POSITIF COVID-19 ≤ 0.500
| NAMA KECAMATAN = DEMPO SELATAN: TIDAK RAWAN {RAWAN=0, TIDAK RAWAN=6}
| NAMA KECAMATAN = DEMPO TENGAH
| | JENIS KELAMIN = L
| | | USIA = ANAK : TIDAK RAWAN {RAWAN=0, TIDAK RAWAN=1}
| | | USIA = TUA: RAWAN {RAWAN=1, TIDAK RAWAN=0}
| | JENIS KELAMIN = P: RAWAN {RAWAN=1, TIDAK RAWAN=0}
| NAMA KECAMATAN = DEMPO UTARA: RAWAN {RAWAN=1, TIDAK RAWAN=0}
| NAMA KECAMATAN = PAGAR ALAM SELATAN: RAWAN {RAWAN=1, TIDAK RAWAN=0}
    
```

Fig. 8. *Rue* pohon keputusan

5. Kesimpulan

Dari hasil penelitian mengenai “Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 Kota Pagar Alam” maka dapat disimpulkan bahwa :

1. Algoritma C4.5 berhasil diterapkan pada klasifikasi data Covid-19 Kota Pagar Alam, dengan memperoleh hasil tingkat akurasi sebesar 86.67% dan memperoleh nilai ROC/AUC 0.989 yang tergolong *Good Classification*.
2. Berdasarkan hasil analisa algoritma C4.5 bahwa pemetaan wilayah kecamatan di kota Pagar Alam hamper seluruhnya dikatkan daerah rawan penyebaran Covid-19, hanya terdapat 1 kecamatan yang dinyatakan tidak rawan dari 5 kecamatan yaitu kecamatan Dempo Selatan.
3. Algoritma C4.5 adalah model klasifikasi yang baik dalam melakukan klasifikasi data covid-19 kota Pagar Alam.
4. Klasifikasi Algoritma C4.5 dengan metode *Operators Cross Validation* dapat meningkatkan kinerja Algoritma C4.5.
5. *Root* atau akar pohon keputusan yang diperoleh dari hitung manual diperoleh

atribut Dampak Covid-19 dengan nilai *gain ratio* 0,287917677, sedangkan hasil pengujian pada *software Rapid Miner* diperoleh atribut Positif Covid-19 sebagai *root* atau akar pohon keputusan.

Adapun saran untuk penelitian selanjutnya untuk meningkatkan hasil yang memuaskan diperlukan penentuan atribut yang lebih banyak dan spesifik agar hasil lebih akurat perlu dilakukan pengkajian dengan menggunakan metode klasifikasi yang lain agar diketahui kinerja algoritma C4.5.

6. Ucapan Terima Kasih

Ucapan terima kasih penulis ucapkan kepada dosen Pembimbing Erwin Dwika Putra, S. Kom, M. Kom, Dinas Kependudukan dan Pencatatan Sipil kota Pagar Alam dan Dinas Kesehatan kota Pagar Alam, serta seluruh pihak yang telah berberan dalam penulisan karya ilmiah ini sehingga dapat selesai dengan baik.

7. Daftar Pustaka

- [1] World Healty Organization, "WHO Director Genera's Opening Remarks At the Media Breating On Covid-19-11 march 2020". 13 November 2020, [Online] Apalable: <https://www.who.int/director-general-s-opening-remarks-at-the-media-breating-on-covid-19-11-march-2020>
- [2] Rafiska R, Defit S, Nurcahyo, G. W. 2018. "Analisis Rekam Medis Untuk Menentukan Pola Kelompok Penyakit." 2(1) : 391 – 96.
- [3] Turnip, S. M, Silitonga, P. 2018. "Analisis Pola Penyebaran Penyakit Dengan Menggunakan Algoritma C4.5." 03(479):3-7.
- [4] Wijaya, Lalu, and Nur Arini Pratiwi. 2020. "Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur." 3(2):64-73.
- [5] Septiani, Wisti Dwi. 2017. "Komprasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis." 13(1):76-84.
- [6] Bastian, Ade, Harun Sujadi, Gigin Febrianto, Program Studi, Teknik Informatika, Universitas Majalengka, Jl Universitas, and Majalengka No. 2018. "Penerapan Algoritma K-Means Clustering Anlysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)." Jurnal Sistem Informasi (Jurnal Of Information System) 14(1):26-32.
- [7] A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," J. Tek. Inform. dan Sist. Inf., vol. 1, pp. 260-270, 2015.

8. Penulis

	Rega Satya Putra adalah mahasiswa Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu, bidang penelitian yang diminati adalah analisis data.
	Erwin Dwika Putra adalah Dosen Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu.
	Muhammad Husni Rifqo adalah Dosen Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu.
	Harry Witriyono adalah Dosen Program Studi Teknik Informatika, Universitas Muhammadiyah Bengkulu.