

Klasifikasi Gejala Awal COVID-19 Dengan Algoritma Classification And Regression Tree (CART)

Agung Alamsyah*, Anita Desiani, Endro Setyo Cahyono

Fakultas MIPA, Universitas Sriwijaya

E-mail Korespondensi : anita_desiani@unsri.ac.id

History Artikel

Diterima : 11 Agustus 2023 Disetujui : 06 September 2023 Dipublikasikan : 20 Oktober 2023

Abstract

COVID-19 is a disease that can cause death and can spread to others. By identifying early symptoms of the disease, early detection can be made for several symptoms that may cause COVID-19. One way to predict COVID-19 is through classification methods. By identifying the symptoms that have an impact on COVID-19, it is hoped that the COVID-19 virus can be stopped from spreading and the world's condition can be normal. This study shows an analysis of attributes that may have an impact on the onset of COVID-19 in an individual. The classification method used is one of the decision tree methods, namely the Classification and Regression Tree (CART). The training and testing methods used in this study are cross-validation and percentage split. The attribute that has a significant influence in this classification using CART method is lung infection. The performance of the system using cross-validation method with a value of k of 10 obtained an accuracy of 85%, which is considered good, while using a percentage split of 66%, an accuracy of 87% was obtained. The evaluation results for the class indicating COVID-19 with precision and recall in cross-validation are 70% and 68%, respectively, while for the percentage split method, precision and recall values of 75% and 70% were obtained, respectively.

Keywords: *COVID-19; Classification; Classification and Regression Tree; Decision Tree*

Abstrak

COVID-19 adalah penyakit yang dapat menyebabkan kematian dan dapat menyebar kepada orang lain. Dengan adanya data gejala awal pada penyakit ini, dapat dilakukan deteksi dini pada beberapa gejala yang dapat menyebabkan penyakit COVID-19. Salah satu cara prediksi penyakit COVID-19 adalah dengan metode klasifikasi. Dengan mengetahui gejala yang memberikan pengaruh COVID-19, diharapkan virus COVID-19 dapat berhenti menyebar dan keadaan dunia dapat membaik. Dalam penelitian ini akan ditunjukkan analisa atribut yang dapat memberikan pengaruh kepada munculnya penyakit COVID-19 pada seseorang. Metode klasifikasi yang digunakan adalah salah satu metode pohon keputusan yaitu Classification and Regression Tree (CART). Metode latihan dan uji yang digunakan dalam penelitian ini yaitu cross validation dan persentase split. Atribut yang memberikan pengaruh besar dalam klasifikasi ini dengan metode CART ialah infeksi paru-paru. Hasil kinerja sistem menggunakan metode cross validation dengan nilai k adalah 10 mendapatkan akurasi sebesar 85% yang terbilang baik, sedangkan menggunakan persentase split 66% didapat akurasi sebesar 87%. Hasil evaluasi pada kelas yang menunjukkan terkena COVID-19 dengan presisi dan recall pada cross validation berturut-turut adalah 70% dan 68%, sedangkan untuk metode persentase split mendapatkan nilai presisi dan recall berturut-turut yaitu 75% dan 70%.

Kata Kunci: COVID-19; Klasifikasi; Classification and Regression Tree; Decision Tree

How to Cite: Alamsyah et. al (2023). Klasifikasi Gejala Awal Covid-19 Dengan Algoritma Classification And Regression Tree (CART). KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo, Vol 7 (2): Halaman 67-76

© 2023 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

PENDAHULUAN

Novel coronavirus yang dikenal sebagai COVID-19 atau 2019-nCoV merupakan penyakit yang dipercaya ditemukan di Wuhan, China pada akhir tahun 2019 (Muhammad et al., 2020). COVID-19 dipercaya berasal dari kelelawar yang dapat menyebar dengan cepat antar tubuh manusia. Gejala umum yang muncul pada penyakit ini ialah batuk, demam, dan kesulitan bernapas dimana gejala ini mirip dengan gejala flu (Muhammad et al., 2020). World Health Organization (WHO) mengkonfirmasi bahwa terdapat 41 kasus dan 1 orang meninggal pada 12 Januari 2020 dan ditetapkan sebagai Public Health Emergency of International Concern (PHEIC) pada 30 Januari 2020 (Li et al., 2020). Semakin berkembangnya kasus COVID-19 dan menyebar ke dalam negeri, data tentang COVID-19 semakin besar dan dibutuhkan klasifikasi pada penyakit tersebut dengan menggunakan gejala-gejala umum seseorang yang terpapar COVID-19 dengan tujuan meminimalisir terjadinya penyebaran virus tersebut.

Salah satu cara untuk memprediksi label pada data gejala awal COVID-19 adalah dengan metode klasifikasi. Salah satu metode non parametrik yang digunakan dalam klasifikasi adalah metode *decision tree*. Salah satu metode *decision tree* yaitu *Classification and Regression Tree* (CART) (Lestawati et al., 2018). Algoritma CART membangun sebuah pohon keputusan biner dengan mengambil *information gain* terbesar di setiap node (Latifah et al., 2019). Algoritma CART

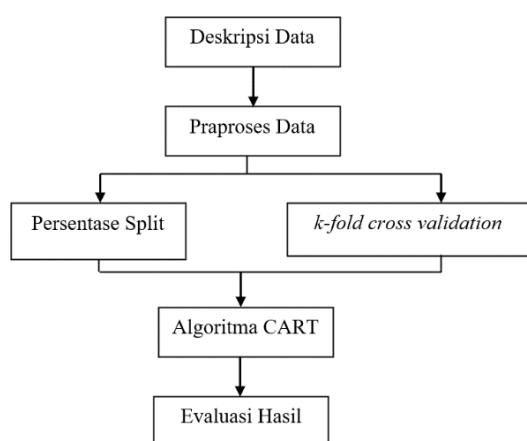
memiliki kelebihan yaitu tidak membutuhkan bentuk fungsional dan menghasilkan pohon keputusan dengan hasil pohon keputusan biner (Monalisa & Hadi, 2020). Algoritma CART juga merupakan klasifikasi yang mudah diinterpretasikan, cepat dalam perhitungannya dan memberikan hasil yang lebih akurat dibandingkan metode lain (Izzati Rahmi & Hazmira Yozza, 2020). Akan tetapi CART memiliki kekurangan yaitu pohon keputusan yang dihasilkan akan menjadi sangat besar sehingga harus dioptimalkan lebih lanjut (Santhanam & Sundaram, 2010). Algoritma CART sudah banyak diaplikasikan untuk bidang kesehatan (Santhanam & Sundaram, 2010). Salah satunya adalah klasifikasi gejala saluran kemih bawah laki-laki dengan hasil presisi 94% (Stothers et al., 2010). Pada tahap pelatihan data akan digunakan metode pembagian data, contohnya yaitu menggunakan *k-fold cross validation* dan persentase split. Metode persentase split adalah pendekatan integral dalam pembelajaran mesin yang menjadi generalisasi pada data (Kebonye, 2021). *K-fold cross validation* adalah metode dengan membagi dataset menjadi *k* bagian dengan jumlah yang sama dan data tersebut akan berlatih secara bergantian (Mileman, 2001).

Pada penelitian ini, akan dilakukan klasifikasi gejala-gejala awal dengan menggunakan metode CART untuk menentukan bahwa seseorang terkena penyakit COVID-19 atau tidak. Data yang digunakan dalam penelitian ini ialah *Early stage symptoms of COVID-19 patient's* dengan 12 atribut, dimana 1 atribut akan menjadi label dan 11 atribut menjadi fitur. Atribut fitur yang digunakan merupakan gejala awal COVID-19 sebagai ciri-ciri

apakah seseorang mengalami gejala-gejala *Early stage symptoms of COVID-19 patient's* yang tersebut atau tidak. Pengujian data akan didapat dari situs kaggle dilakukan dengan dua metode, yaitu *k-fold cross validation* dan persentase split. Hasil Data kinerja kedua metode akan ditunjukkan melalui nilai akurasi, presisi, dan recall. Kedua hasil kinerja tersebut akan dibandingkan untuk mengetahui metode terbaik dalam mengklasifikasikan gejala awal COVID-19 dengan metode CART.

METODE PENELITIAN

Pada penelitian ini, akan dilalui beberapa tahap, yaitu deskripsi data, praproses data, yaitu meliputi pembagian data dengan menggunakan *k-fold cross validation* dan persentase split. Dilanjutkan dengan menerapkan algoritma CART dan diakhiri dengan evaluasi hasil dan pembahasan. Alur metode penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Metode Penelitian

2.1 Deskripsi Data

Sebelum melakukan klasifikasi, perlu dipersiapkan terlebih dahulu dataset yang akan dilakukan klasifikasi dan pengujian. Untuk penelitian ini, dataset yang digunakan bernama

tersebut memuat 12 atribut dan 6512 data. Pada dataset ini, terdapat 11 atribut dan 1 atribut kelas yang menjadi kelas hasil keluaran yang menandakan apakah seseorang dicurigakan terpapar COVID-19 atau tidak. Atribut independen yaitu *gender, age_year, fever, cough, muscle soreness, runny_nose, pneumonia, diarrhea, lung_infection, travel_history,* dan *isolation_treatment*. Sedangkan atribut dependen yang menjadi label yaitu *outcome*. Tidak terdapat data yang hilang pada dataset ini.

Tabel 1. Atribut dan Keterangan Atribut

Atribut	Keterangan
<i>Gender</i>	1=perempuan; 0=laki-laki
<i>Age_year</i>	Usia seseorang
<i>Fever</i>	0 = tidak demam; 1= demam
<i>Cough</i>	0 = tidak batuk; 1= batuk
<i>Runny_nose</i>	0 = hidung tidak tersumbat; 1= hidung tersumbat
<i>Muscle_soreness</i>	0 = tidak nyeri otot; 1= nyeri otot
<i>pneumonia</i>	0 = tidak pneumonia; 1= pneumonia
<i>diarrhea</i>	0 = tidak diare; 1= diare
<i>Lung_infection</i>	0 = tidak infeksi paru-paru; 1= infeksi paru-paru
<i>Travel_history</i>	0 = tidak memiliki riwayat perjalanan; 1= memiliki riwayat perjalanan
<i>Isolation_treatment</i>	0 = tidak menjalankan isolasi; 1= menjalankan isolasi

<i>outcome</i>	0 = Tidak dicurigakan terkena COVID-19; 1= dicurigakan terkena COVID-19
----------------	---

2.2 Praproses Data

Sebelum melakukan klasifikasi, dataset akan dilakukan pemrosesan data. Pada dataset ini akan dilakukan pembagian data menggunakan metode k-fold cross validation dan persentase split. Nilai k yang digunakan pada metode cross validation adalah k=10, yaitu membagi data menjadi 10 bagian, dan melakukan latihan pada salah satu bagian, sedangkan data lainnya akan menjadi data uji, dilakukan secara bergantian hingga seluruh bagian mendapatkan giliran menjadi data latihan. Pada persentase split akan digunakan 66% untuk data latihan dan 34% untuk data uji.

2.3 Implementasi Classification and Regression Tree (CART)

CART membangun pohon keputusan dengan membagi secara rekursif variabel-variabelnya berdasarkan *impurity* hingga memenuhi kondisi tertentu (Santhanam & Sundaram, 2010). Tahapan dalam perhitungan algoritma CART yaitu sebagai berikut (Suntoro, 2019):

1. Menghitung nilai *Gini Index* dalam menentukan cabang pohon keputusan. *Gini Index* dapat dilihat pada persamaan 1:

Dimana k adalah banyaknya indeks, dan p_i^2 adalah peluang sebuah *tuple* pada indeks ke-i.

2. Menghitung nilai *Gini Gain* dalam menentukan *node*. *Gini Gain* dapat dihitung menggunakan persamaan 2:

$$Gini\ Gain = Gini(A, D) - \sum_{i=1}^n \frac{|D_i|}{|D|} \times Gini(D_i) \tag{2}$$

Di mana $Gini(A, D)$ merupakan *impurity* dari suatu partisi A pada atribut D. D_i adalah partisi ke-i pada *tuple* D.

3. Membuat *node* dan cabang dari nilai *Gini Gain* maksimal.
4. Melakukan tahap 1 sampai 3 secara berulang hingga semua *node* sudah terpartisi.

2.4 Evaluasi Hasil

Evaluasi hasil uji kinerja sistem dalam melakukan klasifikasi akan ditampilkan menggunakan confusion matrix dengan ukuran evaluasi kinerja sistem yaitu akurasi, presisi, dan recall. Confusion matrix adalah sebuah informasi yang dihasilkan dari tahap uji mengenai klasifikasi yang sebenarnya dan klasifikasi yang diperkirakan oleh sistem (Bagus & Imron, 2018).

Tabel 2. *Confusion Matrix*

		Prediksi	
		Positif	Negatif
Aktual	Positif	True Positif (TP)	False Negatif (FN)
	Negatif	False Positif (FP)	True Negatif (TN)

$$Gini\ Index = 1 - \sum_{i=1}^k p_i^2 \tag{1}$$

HASIL DAN PEMBAHASAN

Pada tabel 2, hasil *true* positif menandakan klasifikasi diprediksi dengan benar dan bernilai positif. *True* negatif menunjukkan klasifikasi diprediksi benar dan bernilai negatif. Sedangkan yang bernilai *false* positif menunjukkan klasifikasi diprediksi salah dan aktual bernilai negatif. *False* negatif menunjukkan klasifikasi diprediksi salah dan aktual bernilai positif. Hasil evaluasi kinerja sistem yaitu akurasi, presisi dan recall dapat dihitung menggunakan persamaan (Pratiwi et al., 2021):

$$akurasi = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \cdot 100\%$$

$$presisi = \left(\frac{TP}{TP + FP} \right) \cdot 100\%$$

$$recall = \left(\frac{TP}{TP + FN} \right) \cdot 100\%$$

Akurasi adalah ukuran seberapa baik model dapat mengklasifikasikan data dengan benar secara keseluruhan, yaitu rasio antara jumlah prediksi yang benar dengan jumlah total prediksi. Presisi adalah ukuran seberapa baik model dapat mengklasifikasikan data dengan benar untuk kelas positif, yaitu rasio antara jumlah prediksi benar positif dengan jumlah prediksi positif. Recall adalah ukuran seberapa baik model dapat mengidentifikasi dan mengklasifikasikan data dengan benar untuk kelas positif, yaitu rasio antara jumlah prediksi benar positif dengan jumlah data aktual yang benar positif.

Dalam penerapan algoritma CART, perhitungan menggunakan gini gain untuk menentukan pohon keputusan mulai dari node awal hingga node akhir. Atribut *lung_infection* memiliki nilai *gini gain* maksimum pada node awal, sehingga atribut tersebut dipilih sebagai akar pohon keputusan. Hal ini berarti bahwa jika seseorang mengalami infeksi paru-paru, maka kemungkinan besar orang tersebut terinfeksi COVID-19. Namun, jika seseorang tidak mengalami infeksi paru-paru, perlu dilakukan pemeriksaan dengan menggunakan atribut lain yang memiliki nilai *gini gain* maksimum setelah atribut *lung_infection* dipisahkan menjadi nilai 1 dan 0.

- (3) Urutan dari pohon keputusan yang dihasilkan menggunakan algoritma CART yaitu *lung_infection*, *cough*, *pneumonia*, *fever*, *runny_nose*, *travel_history*, dan *isolation_treatment*, *gender*, *age_year*, dan *muscle_soreness*. Untuk analisis secara lengkap keputusan yang didapat, diberikan aturan keputusan untuk klasifikasi gejala awal pada COVID-19 yang ditunjukkan pada gambar 2.

```

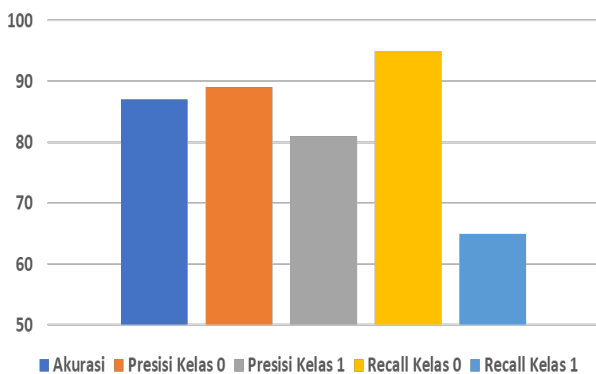
IF lung_infection= 1 THEN outcome= 1
IF lung_infection= 0 AND cough= 0 THEN outcome= 0
IF lung_infection= 0 AND cough= 1 AND pneumonia= 1 THEN outcome= 1
IF lung_infection= 0 AND cough= 1 AND pneumonia= 0 AND fever= 0 THEN outcome= 0
IF lung_infection= 0 AND cough= 1 AND pneumonia= 0 AND fever= 1 AND runny_nose= 1 THEN outcome = 0
IF lung_infection= 0 AND cough= 1 AND pneumonia= 0 AND fever= 1 AND runny_nose= 0 AND travel_history= 0 THEN outcome = 0
IF lung_infection= 0 AND cough= 1 AND pneumonia= 0 AND fever= 1 AND runny_nose= 0 AND travel_history= 1 AND isolation_treatment=1 THEN outcome= 1
IF lung_infection= 0 AND cough= 1 AND pneumonia= 0 AND fever= 1 AND runny_nose= 0 AND travel_history= 1 AND isolation_treatment=0 THEN outcome= 0
    
```

Gambar 2. Aturan Keputusan

Pada Gambar 2, setelah aturan linguistik mengenai keputusan telah terbentuk, sistem akan diuji untuk mengevaluasi keberhasilan model yang telah dibuat. Evaluasi hasil dilakukan menggunakan *confusion matrix* dan menghitung akurasi, *recall*, dan presisi untuk mengukur kinerja sistem.

3.1 Persentase Split

Dengan metode latihan dengan persentase split 66%, data akan dibagi sebanyak 66% dari total data menjadi data latihan, yaitu sebanyak 4298 data, sedangkan data uji digunakan sebanyak 34% atau sebanyak 2214. Setelah dilakukan pengujian didapat sebuah *confusion matrix* dan hasil evaluasi kinerja sistem. Hasil secara keseluruhan dari evaluasi kinerja sistem ditunjukkan pada gambar 3.



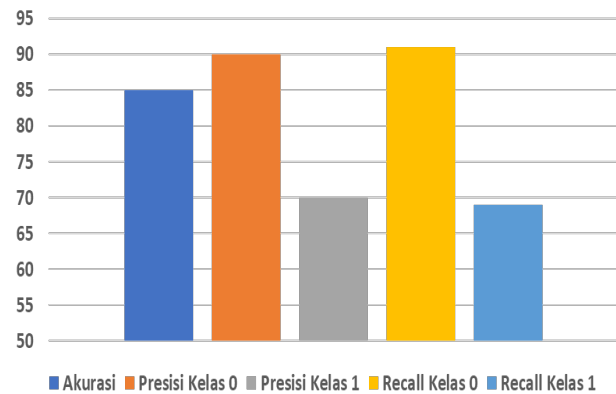
Gambar 3. Evaluasi Hasil Persentase Split 66%

Hasil evaluasi pada Gambar 3 menunjukkan nilai akurasi sebesar 87%, dengan *recall* untuk kelas 1 sebesar 70% dan untuk kelas 0 sebesar 93%. Presisi untuk kelas 1 adalah 75%, sedangkan untuk kelas 0 adalah 91%. Dalam hal rata-rata presisi, nilai yang dicapai adalah sebesar 83%, sementara *recall* mendapatkan hasil rata-rata 81%. Untuk

evaluasi hasil kinerja sistem, didapat nilai akurasi yang terbilang cukup baik yaitu 87%. Akan tetapi hasil presisi dan *recall* untuk kelas 1 menghasilkan nilai yang cukup kecil.

3.2 K-fold Cross Validation

Pada metode *k-fold cross validation* akan digunakan $k=10$, yaitu data dibagi menjadi 10 bagian yang dilakukan latihan pada satu bagian dan uji pada data lainnya dan dilakukan secara bergantian. Hasil sistem ditunjukkan pada gambar 4.



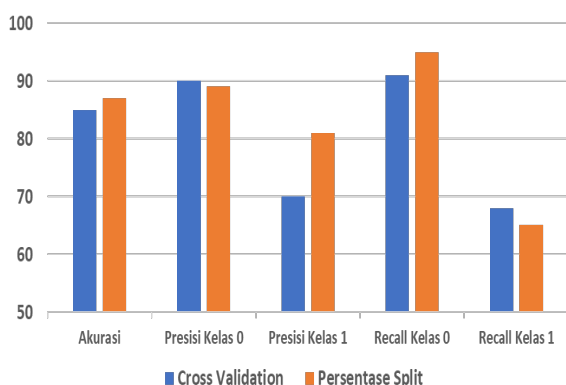
Gambar 4. Evaluasi Hasil K-Fold Cross Validation

Pada gambar 4, hasil evaluasi menunjukkan nilai akurasi sebesar 85%, dengan *recall* untuk kelas 1 sebesar 68% dan untuk kelas 0 sebesar 91%. Presisi untuk kelas 1 adalah 70%, sedangkan untuk kelas 0 adalah 90%. Dalam hal rata-rata presisi, nilai yang dicapai adalah sebesar 80%, sementara *recall* mendapatkan hasil rata-rata 79%. Untuk evaluasi hasil kinerja sistem, didapat nilai akurasi yang terbilang cukup baik yaitu 85%. Tetapi hasil presisi dan *recall* untuk kelas 1 menghasilkan nilai yang kecil seperti pada persentase split. Hal ini dapat disebabkan oleh jumlah data yang tidak seimbang. Pada dataset ini,

data dengan kelas 1 lebih sedikit dibandingkan dengan data kelas 0. Sehingga hasil presisi maupun recall pada kedua implementasi tersebut tergolong cukup kecil untuk kelas 1.

3.3 Perbandingan Metode Persentase Split dan K-Fold Cross Validation

Pada metode persentase split yang digunakan adalah 66% data sebagai data latihan dan 34% sebagai data uji. Sedangkan pada metode k-fold cross validation dengan menggunakan nilai k yaitu 10, seluruh data mendapat bagian untuk menjadi data latih dan data uji. Perbandingan dari kedua metode disajikan pada gambar 5.



Gambar 5. Perbandingan Metode K-Fold Cross Validation dan Persentase Split

Pada gambar 5 dapat dilihat bahwa evaluasi hasil pada metode CART dengan metode k-fold cross validation dan persentase split didapat hasil yang baik dan hanya memiliki perbedaan yang cukup signifikan untuk presisi dan recall kelas 1 dan juga recall kelas 0. Kedua metode latihan dan uji yang digunakan memberikan nilai akurasi, presisi dan recall

diatas 65%. Hasil presisi dan recall diperoleh dengan nilai yang kecil untuk kelas 1 dengan kedua metode karena jumlah data dengan hasil keluaran kelas 1 lebih sedikit dibandingkan jumlah data hasil keluaran kelas 0 yang terdapat pada dataset tersebut. Secara keseluruhan, persentase split dengan menggunakan 66% data training dan 34% data testing dapat memberikan hasil yang lebih baik dibandingkan k-fold cross validation dengan nilai k yaitu 10.

3.4 Perbandingan Algoritma CART dengan Algoritma Lainnya

Perbandingan metode CART dengan metode yang sudah dilakukan oleh peneliti sebelumnya dengan data *Early stage symptoms of COVID-19 patient's* ditunjukkan pada tabel 3.

Tabel 3. Perbandingan Algoritma CART dengan Algoritma Lain

Nama Peneliti	Algoritma	Presisi	Recall	Akurasi
Agung Alamsyah (Ahamad et al., 2020)	CART	83%	81%	87%
	XGBoost	93%	91%	-
	GBM	97%	85%	-
	SVM	97%	84%	-
	Random Forest	92%	89%	-
	Decision Tree	91%	90%	-

Tabel 3 menunjukkan bahwa algoritma CART menggunakan metode persentase split menghasilkan rata-rata presisi sebesar 83% dan recall sebesar 81%. Penelitian ini menunjukkan bahwa metode CART memberikan hasil yang lebih baik ketika menggunakan persentase split sebesar

66%, dengan akurasi, presisi, dan recall yang lebih baik dibandingkan dengan metode *k-fold cross validation*. Secara keseluruhan, kinerja sistem dengan metode CART tergolong baik dan kinerja hasil dapat menjadi lebih baik jika jumlah data antar kelas seimbang dan data yang digunakan bersifat kategorikal.

KESIMPULAN

Berdasarkan hasil penelitian, algoritma CART dengan metode persentase split 66% menghasilkan akurasi sebesar 87%, yang dapat dikategorikan sebagai tingkat keakuratan yang baik. Hasil ini juga dapat dibandingkan dengan studi sebelumnya yang menggunakan dataset yang sama, dan CART menunjukkan hasil yang cukup baik dalam hal presisi, *recall*, dan akurasi. Pada evaluasi hasil, metode persentase split sebesar 66% menghasilkan performa yang lebih baik dibandingkan metode *k-fold cross validation*, terutama dalam hal akurasi, presisi, dan recall. Dalam penerapan metode pohon keputusan CART, atribut pertama yang menjadi akar adalah infeksi paru-paru, sehingga jika seseorang mengalami infeksi paru-paru, maka kemungkinan terinfeksi COVID-19 juga tinggi. Namun, jika seseorang tidak mengalami infeksi paru-paru, atribut lain yang memiliki nilai gini gain tertinggi harus dievaluasi hingga kondisi tertentu terpenuhi sehingga bisa diambil keputusan secara tepat. Oleh karena itu, dengan pengetahuan tentang algoritma CART dalam memprediksi gejala awal COVID-19, diharapkan dapat membantu masyarakat dalam mendeteksi dini penyakit ini dan dapat

mencegah penyebarannya.

Saran untuk penelitian selanjutnya adalah pengumpulan dataset bisa lebih banyak dengan menyeimbangkan jumlah data per kelasnya. Untuk permasalahan kekurangan data yang ada pada kelas tertentu, bisa digunakan metode statistika dalam menangani data yang tidak seimbang. Pemilihan metode pembagian data bisa disesuaikan dengan kebutuhan penelitian dan data yang tersedia.

REFERENSI

- Ahamad, M. M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H., Summers, M. A., Quinn, J. M. W., & Moni, M. A. (2020). A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. *Expert systems with applications*, *160*, 113661.
- Bagus, C., & Imron, M. (2018). Klasifikasi Buah Mangga Berdasarkan Tingkat Kematangan Menggunakan Least-Squares Support Vector Machine. *Explore IT : Jurnal Keilmuan dan Aplikasi Teknik Informatika*, *10*(2), 1–8. <https://doi.org/10.35891/explorit.v10i2.1255>
- Izzati Rahmi, & Hazmira Yozza. (2020). Analisis Kausal Masa Studi Mahasiswa Program Studi Matematika Universitas Andalas Dengan Menggunakan Metode Cart. *Jurnal Matematika Sains dan Teknologi*, *21*(1), 22–34. <https://doi.org/10.33830/jmst.v21i1.697.2020>
- Kebonye, N. M. (2021). Exploring the novel support points-based split method on a soil dataset. *Measurement: Journal of the International Measurement Confederation*, *186*, 110131. <https://doi.org/10.1016/j.measurement.2021.110131>
- Latifah, R., Wulandari, E. S., & Kreshna, P. E. (2019). Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn. *Jurnal Universitas Muhammadiyah Jakarta*, 1–6. <https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517>
- Lestawati, R., Rais, R., & Utami, I. T. (2018). Perbandingan Antara Metode Cart (Classification and Egression Tree) Dan

- Regresi Logistik (Logistic Regression) Dalam Mengklasifikasikan Pasien Penderita Dbd (Demam Berdarah Dengue). *Jurnal Ilmiah Matematika Dan Terapan*, 15(1), 98–107.
- Li, J., Xu, Q., Cuomo, R., Purushothaman, V., & Mackey, T. (2020). Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Inveillance Study. *JMIR Public Health Surveill*, 6(2), e18700. <https://doi.org/10.2196/18700>
- Mileman, P. A. (2001). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence (IJCAI)*, 30(2), 133–133.
- Monalisa, S., & Hadi, F. (2020). Penerapan Algoritma CART Dalam Menentukan Jurusan Siswa di MAN 1 Inhil. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 9(3), 387–394. <https://doi.org/10.32736/sisfokom.v9i3.932>
- Muhammad, L. J., Islam, M. M., Usman, S. S., & Ayon, S. I. (2020). Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN Computer Science*, 1(4), 1–7. <https://doi.org/10.1007/s42979-020-00216-w>
- Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, 6(2), 66–75. <https://doi.org/10.26877/jiu.v6i2.6552>
- Santhanam, T., & Sundaram, S. (2010). Application of CART algorithm in blood donors classification. *Journal of Computer Science*, 6(5), 548–552. <https://doi.org/10.3844/jcssp.2010.548.552>
- Stothers, L., Guevara, R., & Macnab, A. (2010). Classification of Male Lower Urinary Tract Symptoms Using Mathematical Modelling and a Regression Tree Algorithm of Noninvasive Near-Infrared Spectroscopy Parameters. *European Urology*, 57(2), 327–333. <https://doi.org/https://doi.org/10.1016/j.eururo.2009.05.004>
- Suntoro, J. (2019). Data Mining Algoritma dan Implementasi dengan Pemrograman PHP. In *Elex Media Komputindo* (Vol. 1). Elex Media Komputindo.