

Implementasi Algoritma C4.5 untuk Klasifikasi Tingkat Kenaikan Kasus Covid-19

Bari Anandanu Wibowo¹, Clarissa Nabila², Dini Nursafitri³
Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pondok Labu, Jakarta Selatan 12450

bari.anandanu99@gmail.com^{1*}, clarissanabila@gmail.com^{2*}, dininursafitri30@gmail.com^{3*}

Abstrak. *Coronavirus Disease* (Covid-19) merupakan penyakit menular yang disebabkan oleh virus corona yang pertama kali terdeteksi pada tahun 2019 di Wuhan, China. Indonesia termasuk salah satu negara yang masih berkebutakan mengempur virus Corona sampai saat ini. Berbagai cara sudah dilakukan oleh Pemerintah dalam mengantisipasi dan menanggulangi pandemi ini, seperti memberlakukan kebijakan *Work from Home*, PSBB, dan lainnya. Penelitian ini dilakukan sebagai langkah untukantisipasi masyarakat terhadap Covid-19 dengan mengelompokkan tingkat kenaikan kasus virus corona yang terjadi di beberapa kota di Indonesia. Pada penelitian ini peneliti menggunakan penerapan algoritma C4.5 diharapkan dapat mengelompokkan tingkat kenaikan kasus Covid-19 agar masyarakat mengetahui tingkat kenaikan kasus di daerahnya dan lebih *aware* dalam menjaga kesehatan, serta selalu mematuhi protokol kesehatan. Hasil penelitian pengelompokan tingkat kenaikan kasus Covid-19 menggunakan algoritma C4.5 ini menunjukkan bahwa didapat keakuratan sebesar 83,33%.

Kata Kunci: Covid-19, Algoritma C4.5, pengelompokan

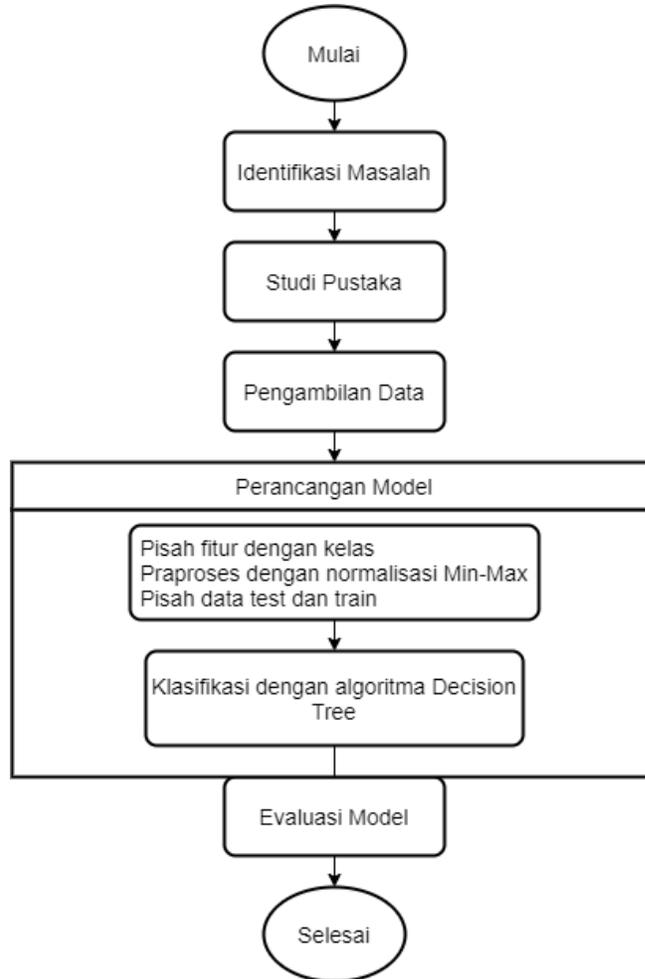
1 Pendahuluan

Indonesia beserta dengan negara lain yang ada di dunia masih berkebutakan mengempur virus Corona sampai saat ini. Seiring dengan berjalannya waktu, jumlah kasus virus Corona semakin bertambah [1], baik itu kasus positif, kesembuhan, dalam perawatan maupun kasus meninggal. Penyakit Coronavirus (Covid-19) adalah penyakit menular yang disebabkan oleh virus corona yang baru ditemukan [2]. Virus ini muncul pertama kali tahun 2019, di provinsi Wuhan, China. Kasus pada tahun 2019 tersebut terus berlanjut hingga penyebaran virus Corona mewabah sampai ke seluruh dunia. Karena penularannya yang sangat cepat, pada tanggal 11 Maret 2020 WHO (*World Health Organization*) menetapkan virus Corona sebagai pandemi.

Mengutip dari “Worldometers”, di Indonesia per tanggal 28 Maret 2021 terkonfirmasi kasus Covid-19 sebanyak 1,496,085 juta kasus [3], dengan provinsi DKI Jakarta berada di posisi pertama, lalu disusul oleh Jawa Barat dan Jawa Timur. Berdasarkan kondisi tersebut, Indonesia masuk kedalam status waspada terhadap ancaman Covid-19. Oleh karena itu, untuk mencegah penyebaran Covid-19 dapat dilakukan dengan melakukan pembatasan social atau *social distancing*. Dengan adanya pembatasan social tersebut, diharapkan masyarakat tidak akan menjadi tertular ataupun penular, sehingga lajur penyebaran Covid-19 dapat menurun. Pada penelitian ini, peneliti melakukan pengelompokkan tingkat kenaikan kasus Covid-19 untuk mengetahui tingkat kenaikan kasus di beberapa daerah di Indonesia, sehingga masyarakat menjadi lebih waspada dalam menjaga kesehatan dan mematuhi protocol kesehatan. Pada penelitian sebelumnya yang dilangsungkan oleh Watratan dkk., mereka telah melakukan pengelompokan tingkat penyebaran virus Covid-19 di Indonesia menjadi 3 kelompok, yaitu rendah, sedang, dan tinggi menggunakan algoritma *Naive Bayes* dan memperoleh hasil akurasi sebesar 48,4848% dengan jumlah 16 dari 33 data yang diujikan berhasil dikelompokkan secara benar [4]. Selain algoritma *Naive Bayes*, terdapat juga algoritma klasifikasi yang lain seperti algoritma C4.5. Algoritma C4.5 termasuk salah satu algoritma klasifikasi yang digunakan untuk membuat pohon keputusan (*decision tree*). Tujuan dari pohon keputusan antara lain, mengubah tabel data menjadi model pohon, lalu mengubah model pohon tersebut dan menyederhanakan aturannya [5]. Dalam penelitian ini algoritma C4.5 diperlukan untuk mengelompokkan tingkatan dari risiko kenaikan kasus Covid-19 menggunakan dataset dengan fitur berupa total keseluruhan pasien Covid-19, jumlah pasien yang sembuh, jumlah pasien yang meninggal, serta jumlah pasien yang masih dalam perawatan. Oleh karena itu, pada penelitian ini peneliti melakukan pengelompokan tingkat kenaikan kasus Covid-19 menjadi 3 kelompok, yaitu rendah, sedang, dan tinggi dengan menggunakan algoritma C4.5 yang diharapkan dapat meningkatkan akurasi ketepatan dalam mengelompokkan data dibandingkan dengan algoritma *Naive Bayes*.

Luaran yang dihasilkan dari penelitian ini adalah pengelompokan dataset Covid-19 dengan 3 kelompok tingkatan dari risiko kenaikan kasus Covid-19, yaitu rendah, sedang, dan tinggi, serta akurasi dan performa dari algoritma C4.5.

2 Metodologi Penelitian



Gambar 1. Metodologi Penelitian

Terdapat metodologi penelitian yang dapat dilihat pada gambar 1 yang akan dijelaskan sebagai berikut:

2.1 Identifikasi Masalah

Identifikasi masalah dilakukan untuk menemukan permasalahan yang ada terutama terhadap kenaikan kasus Covid-19 di beberapa daerah di Indonesia. Pada saat ini beberapa wilayah di Indonesia mengalami kenaikan jumlah kasus Covid-19, untuk mencegah semakin meningkatnya kasus, maka diperlukan kelompok tingkat kenaikan kasus Covid-19, yaitu rendah, sedang, dan tinggi. Dengan mengetahui kelompok kasus Covid-19 ini diharapkan dapat meningkatkan kewaspadaan masyarakat terhadap virus Covid-19 sehingga tertib dalam menjalankan protokol kesehatan dan mengurangi laju penyebaran virus di lingkungan masyarakat.

2.2 Studi Pustaka

Studi pustaka dilakukan oleh peneliti dengan tujuan untuk mempelajari penelitian sebelumnya yang berkaitan dengan penelitian yang dibuat oleh peneliti saat ini. Dalam hal ini, peneliti mempelajari mengenai data tingkat kenaikan kasus Covid-19 di Indonesia, metode dari algoritma C4.5, serta informasi-informasi lainnya yang berkaitan dengan

penelitian yang dilakukan oleh peneliti. Dalam penelitian ini digunakan algoritma C4.5, yaitu salah satu jenis klasifikasi yang merepresentasikan nilai atribut dari presentasi kelasnya [6] dengan langkah-langkah sebagai berikut:

1. Pilih atribut sebagai root.
2. Buat cabang dari masing-masing nilai dari atributnya.
3. Kelompokkan kasus dalam simpulnya.

Root node terletak paling atas dari suatu pohon keputusan. Terdapat perhitungan yang dapat menampilkan *root node* sebagai berikut:

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

S adalah melambangkan kasus, A melambangkan fitur, dan n melambangkan total contoh A ke partisi I dilambangkan $|S_i|$.

$$Entropy(S) = \sum_{i=1}^n -P_i \cdot p_i \quad (2)$$

Perhitungan entropy dilakukan untuk mencari nilai atribut untuk cabang-cabangnya, dengan n jumlah atribut dari S dan p_i proporsi dari nilai S_i .

2.3 Pengambilan Data

Data yang digunakan dalam penelitian ini adalah dataset tingkat kenaikan kasus Covid-19 [7] di beberapa provinsi di Indonesia yang berjumlah 74 *record* yang diperoleh dari beberapa *web* pemerintah di Indonesia yang menyediakan data Covid-19 per tanggal 28 Maret 2021 dan 04 April 2021, yaitu Jawa Barat [8], Bali [9], Kalimantan Tengah [10], Kabupaten Kupang [11], Kapuas [12], Kota Tangerang Selatan [13], dan Kota Mataram [14].

2.4 Perancangan Model

Perancangan model dimulai dari memasukkan dataset Covid-19 dengan fitur berupa total keseluruhan pasien Covid19, jumlah pasien yang sembuh, jumlah pasien yang meninggal, dan jumlah pasien yang masih dalam perawatan serta 3 kelompok yang dijadikan kelas, yakni rendah, sedang, dan tinggi. Lalu dilakukan pemisahan data fitur dengan kelas, yang kemudian dilanjutkan dengan tahapan praproses data fitur menggunakan teknik normalisasi *min-max*. Teknik normalisasi *min-max* ini memiliki tujuan membentuk fitur memiliki rentang nilai yang serupa agar adanya keseimbangan nilai antara data sebelum dan setelah normalisasi, sehingga tidak ada data yang bias dengan cara nilai pada suatu fitur dikurangi dengan nilai terkecil dari fitur tersebut lalu dibagi dengan nilai terbesar dikurangi nilai terkecil dari fitur tersebut. Rumus dari normalisasi *min-max* dapat dilihat sebagai berikut [15]:

$$\text{nilai baru} = \frac{\text{nilai lama} - \text{nilai min}}{\text{nilai maks} - \text{nilai min}} (\text{maks baru} - \text{min baru}) + \text{min baru} \quad (3)$$

Setelah itu, membagikan data dengan menggunakan *3-fold* untuk mendapatkan data *train* yang dipakai untuk pelatihan model dan data *test* untuk melakukan pengujian pengelompokan dengan model. Kemudian latih data *train* dengan model *decision tree* (C4.5) agar dapat mengelompokkan data *test*. Langkah terakhir adalah mengevaluasi hasil pengelompokan *data test* dengan membuat matriks konfusi, menghitung akurasi, presisi, *Recall*, dan *f1-score*.

2.5 Evaluasi Model

Evaluasi model dilakukan dalam penelitian ini dengan menerapkan perhitungan matriks konfusi yang meliputi perhitungan akurasi, *sensitivity*, dan *specificity* untuk mengetahui ketepatan dalam mengelompokkan data dari model yang dibuat terhadap data. Penjelasan dari matriks konfusi, akurasi, presisi, dan *specificity* akan dijelaskan sebagai berikut [16]:

- a. Matriks konfusi merupakan evaluasi performa dari model pengelompokan dimana *output* bisa berupa 2 kelompok atau lebih yang mengidentifikasi prediksi dan kondisi yang sebenarnya dari data.
- b. Akurasi merupakan rasio perkiraan seberapa akurat model dalam mengelompokkan keseluruhan data dengan tepat.

$$Akurasi = \frac{(True\ Positive + True\ Negative)}{(True\ Positive + True\ Negative + False\ Positive + False\ Negative)} \quad (4)$$

- c. *Sensitivity* merupakan pengujian untuk mengukur seberapa baik model dalam mengelompokkan data positif. Pada penelitian ini mengasumsikan bahwa kelompok rendah merupakan kelompok positif, sehingga *sensitivity* mengukur seberapa akurat data kelompok positif yang sebenarnya diprediksi dengan tepat sebagai kelompok positif.

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (5)$$

- d. *Specificity* merupakan pengujian untuk mengukur seberapa baik model dalam mengelompokkan data negatif. Pada penelitian ini mengasumsikan bahwa kelompok sedang dan tinggi merupakan kelompok negatif, sehingga *specificity* mengukur seberapa akurat data kelompok negatif yang sebenarnya diprediksi dengan tepat sebagai kelompok negatif.

$$Specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \quad (6)$$

2.6 Kesimpulan dan Saran

Tahap akhir dari penelitian ini yaitu hasil yang didapat dari menjalankan proses secara keseluruhan akan ditarik kesimpulannya dan diberi saran untuk penelitian kedepannya.

3 Hasil dan Pembahasan

3.1 Analisis Deskriptif Dataset Tingkat Kenaikan Kasus Covid-19

Dataset tingkat kenaikan kasus Covid-19 memiliki 4 fitur, yaitu 'total', 'sembuh', 'meninggal', dan 'isolasi', serta memiliki kelas 'tingkat kenaikan'. Jumlah *record dataset* berjumlah 74 data. Dalam kelas 'tingkat kenaikan' mempunyai 3 macam nilai kelompok yakni 'rendah', 'sedang', dan 'tinggi'. Dari data tersebut yang digunakan untuk sebagai fitur penentu dalam pengolahan dan pengelompokan data adalah fitur 'total', 'sembuh', 'meninggal', dan 'isolasi'.

3.2 Analisis Hasil pengelompokan Tingkat Kenaikan kasus Covid-19

Pada dataset tingkat kenaikan kasus Covid-19 yang telah di-*preprocess* dengan normalisasi metode *Min-Max*, serta pembagian data dengan metode *K-Fold Cross Validation* yang memiliki nilai $K=3$, yang artinya akan dibagi rata menjadi 3 data *train* dan data *test* kemudian dikelompokkan menggunakan algoritma C4.5 dengan fungsi kernel linear. Kemudian dilakukan perhitungan matriks konfusi dengan asumsi kelompok rendah merupakan kelompok positif, sedangkan kelompok sedang dan tinggi merupakan kelompok negatif dari setiap fold yang didapatkan hasil sebagai berikut:

Tabel 1. Matriks Konfusi 3x3

		Hasil Prediksi		
		Rendah	Sedang	Tinggi
Hasil yang Diharapkan	Rendah	True Positive	False Negative	
	Sedang	False Positive	True Negative	
	Tinggi			

Tabel 2. Matriks Konfusi Fold ke-1

Fold Ke-1		Hasil Prediksi		
		Rendah	Sedang	Tinggi
Hasil yang Diharapkan	Rendah	2	2	0
	Sedang	2	10	4
	Tinggi	0	1	4

Pada tabel 2 matriks konfusi menunjukkan bahwa terdapat 2 data kelas “Rendah” yang diprediksi benar sebagai kelas “Rendah”, 2 data kelas “Rendah” yang diprediksi salah sebagai kelas “Sedang”, 0 data kelas “Rendah” yang diprediksi salah sebagai kelas “Tinggi”. Selanjutnya, 2 data kelas “Sedang” yang diprediksi salah sebagai kelas “Rendah”, 10 data kelas “Sedang” yang diprediksi benar sebagai kelas “Sedang”, 4 data kelas “Sedang” yang diprediksi salah sebagai kelas “Tinggi”. Selanjutnya, 0 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Rendah”, 1 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Sedang”, dan 4 data kelas “Tinggi” yang diprediksi benar sebagai kelas “Tinggi”.

Pada *fold* ke-1 ini dilakukan uji performa, sebagai berikut:

$$\text{Akurasi} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$$

$$\text{Akurasi} = \frac{(2 + (10 + 4 + 1 + 4))}{(2 + (10 + 4 + 1 + 4) + (2 + 0) + (2 + 0))} = \frac{21}{25} = 0,84 * 100\% = 84\%$$

Akurasi yang didapat dalam pengelompokkan menggunakan model C4.5 adalah 84%.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{2}{2 + (2 + 0)} = \frac{2}{4} = 0,50 * 100\% = 50\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok positif (rendah) dengan benar sebesar 50%.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} = \frac{(10 + 4 + 1 + 4)}{(2 + 0) + (10 + 4 + 1 + 4)} = \frac{19}{21} = 0,9048 * 100\% = 90,48\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok negatif (sedang dan tinggi) dengan benar sebesar 90,48%.

Tabel 3. Matriks Konfusi Fold ke-2

Fold Ke-2		Hasil Prediksi		
		Rendah	Sedang	Tinggi
Hasil yang Diharapkan	Rendah	2	3	0
	Sedang	2	11	2
	Tinggi	0	4	1

Pada tabel 3 matriks konfusi menunjukkan bahwa terdapat 2 data kelas “Rendah” yang diprediksi benar sebagai kelas “Rendah”, 3 data kelas “Rendah” yang diprediksi salah sebagai kelas “Sedang”, 0 data kelas “Rendah” yang diprediksi salah sebagai kelas “Tinggi”. Kemudian 2 data kelas “Sedang” yang diprediksi salah sebagai kelas “Rendah”, 11 data kelas “Sedang” yang diprediksi benar sebagai kelas “Sedang”, 2 data kelas “Sedang” yang diprediksi salah sebagai kelas “Tinggi”, dan 1 data kelas “Tinggi” yang diprediksi benar sebagai kelas “Tinggi”.

kelas “Tinggi”. Sedangkan 0 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Rendah”, 4 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Sedang”, dan 1 data kelas “Tinggi” yang diprediksi benar sebagai kelas “Tinggi”.

Pada *fold* ke-2 ini dilakukan uji performa, sebagai berikut:

$$\text{Akurasi} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$$

$$\text{Akurasi} = \frac{(2 + (11 + 2 + 4 + 1))}{(2 + (11 + 2 + 4 + 1) + (2 + 0) + (3 + 0))} = \frac{20}{25} = 0,80 * 100\% = 80\%$$

Akurasi yang didapat dalam pengelompokkan menggunakan model C4.5 adalah 80%.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{2}{2 + (3 + 0)} = \frac{2}{5} = 0,40 * 100\% = 40\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok positif (rendah) dengan benar sebesar 40%.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} = \frac{(11 + 2 + 4 + 1)}{(2 + 0) + (11 + 2 + 4 + 1)} = \frac{19}{20} = 0,95 * 100\% = 95\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok negatif (sedang dan tinggi) dengan benar sebesar 95%.

Tabel 4. Matriks Konfusi Fold ke-3

Fold Ke-3		Hasil Prediksi		
		Rendah	Sedang	Tinggi
Hasil yang Diharapkan	Rendah	4	0	0
	Sedang	3	5	7
	Tinggi	0	1	4

Pada tabel 4 matriks konfusi menunjukkan bahwa terdapat 4 data kelas “Rendah” yang diprediksi benar sebagai kelas “Rendah”, 0 data kelas “Rendah” yang diprediksi salah sebagai kelas “Sedang”, 0 data kelas “Rendah” yang diprediksi salah sebagai kelas “Tinggi”. Kemudian 3 data kelas “Sedang” yang diprediksi salah sebagai kelas “Rendah”, 5 data kelas “Sedang” yang diprediksi benar sebagai kelas “Sedang”, 7 data kelas “Sedang” yang diprediksi salah sebagai kelas “Tinggi”. Sedangkan 0 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Rendah”, 1 data kelas “Tinggi” yang diprediksi salah sebagai kelas “Sedang”, dan 4 data kelas “Tinggi” yang diprediksi benar sebagai kelas “Tinggi”.

Pada *fold* ke-3 ini dilakukan uji performa, sebagai berikut:

$$\text{Akurasi} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$$

$$\text{Akurasi} = \frac{(4 + (5 + 7 + 1 + 4))}{(4 + (5 + 7 + 1 + 4) + (3 + 0) + (0 + 0))} = \frac{21}{24} = 0,875 * 100\% = 87,5\%$$

Akurasi yang didapat dalam pengelompokkan menggunakan model C4.5 adalah 87,5%.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{4}{4 + (0 + 0)} = \frac{4}{4} = 1 * 100\% = 100\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok positif (rendah) dengan benar sebesar 100%.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} = \frac{(5 + 7 + 1 + 4)}{(3 + 0) + (5 + 7 + 1 + 4)} = \frac{17}{20} = 0,85 * 100\% = 85\%$$

Tingkat keakuratan model dalam mengelompokkan kelompok negatif (sedang dan tinggi) dengan benar sebesar 85%.

Berdasarkan perhitungan akurasi, presisi, dan *recall* dari masing-masing kelas, maka didapatkan rata-rata akurasi, presisi, dan *recall* sebagai berikut:

$$\text{Rata - rata Akurasi} = \frac{84\% + 80\% + 87,5\%}{3} = \frac{251,5}{3} = 83,33\%$$

$$\text{Rata - rata Sensitivity} = \frac{50\% + 40\% + 100\%}{3} = \frac{190}{3} = 63,33\%$$

$$\text{Rata - rata Specificity} = \frac{90,48\% + 95\% + 85\%}{3} = \frac{270,48}{3} = 90,16\%$$

Berdasarkan hasil klasifikasi dan pengujian yang diperoleh dari penjelasan di atas bahwa menggunakan metode C4.5 memperoleh rata-rata akurasi berdasarkan matriks konfusi yang dihasilkan dengan kelompok rendah merupakan kelompok positif sedangkan kelompok sedang dan tinggi merupakan kelompok negatif adalah sebesar 83,33%. Untuk nilai rata-rata *sensitivity* menunjukkan hasil sebesar 63,33% tingkat keakuratan model dalam mengelompokkan kelompok positif (rendah) dengan benar. Untuk nilai rata-rata *specificity* menunjukkan hasil sebesar 90,16% tingkat keakuratan model dalam mengelompokkan kelompok negatif (sedang dan tinggi) dengan benar.

4 Penutup

4.1 Kesimpulan

Berdasarkan hasil pembahasan menggunakan metode C4.5 dalam proses klasifikasi tingkat kenaikan kasus Covid-19 yang telah dilakukan, maka dapat diperoleh kesimpulan sebagai berikut:

1. Hasil yang didapat dari evaluasi model dengan metode C4.5 berdasarkan matriks konfusi yang dihasilkan dengan kelompok rendah merupakan kelompok positif sedangkan kelompok sedang dan tinggi merupakan kelompok negatif, yaitu nilai akurasi, *sensitivity*, dan *specificity* yang bertujuan untuk mengetahui seberapa baik model dalam mengelompokkan data. Dari hasil evaluasi model pada pembahasan, maka didapatkan nilai rata-rata *sensitivity* sebesar 63,33% dalam mengelompokkan kelompok positif (rendah) dengan benar. Sedangkan, nilai rata-rata *specificity* menunjukkan hasil sebesar 90,16% tingkat keakuratan model dalam mengelompokkan kelompok negatif (sedang dan tinggi) dengan benar.
2. Rata-rata akurasi keseluruhan kelas didapatkan sebesar 83,33%, hal tersebut menunjukkan bahwa pengelompokan tingkat kenaikan kasus Covid-19 dengan metode C4.5 sudah cukup optimal dalam proses pengelompokkan datanya.

4.2 Saran

Berdasarkan analisis yang telah dilakukan, maka peneliti memberikan beberapa saran sebagai berikut:

1. Perlu diuji coba dengan menggunakan dataset Covid-19 yang lebih banyak lagi, hal ini dilakukan untuk melatih data yang lebih banyak dengan algoritma C4.5 sehingga memperoleh nilai akurasi yang lebih akurat.
2. Untuk penelitian kedepannya mengenai pengujian pengelompokkan tingkat kenaikan kasus Covid-19, sebaiknya menggunakan metode lain untuk mendapatkan perbandingan metode mana yang menghasilkan tingkat akurasi yang lebih tinggi.

Referensi

- [1] "Analisis Data COVID-19 Indonesia," *Satuan Tugas Penanganan COVID-19*, 2021. <https://covid19.go.id/p/berita/analisis-data-covid-19-indonesia-update-04-april-2021> (diakses Apr 13, 2021).
- [2] "Coronavirus," *WHO*. https://www.who.int/health-topics/coronavirus#tab=tab_1 (diakses Feb 26, 2021).
- [3] "Coronavirus," *worldometer*. <https://www.worldometers.info/coronavirus/> (diakses Mar 31, 2021).

- [4] A. F. Watratan, A. Puspita, dan D. Moeis, “Implementasi Algoritma Naive Bayes untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia,” *J. Appl. Comput. Sci. Technol.*, vol. 1, hal. 7–14, 2020.
- [5] J. A. Suyatno, F. Nhita, dan A. A. Rohmawati, “Rainfall Forecasting in Bandung Regency using C4.5 Algorithm,” *IEEE Int. Conf. Inf. Commun. Technol.*, hal. 324–328, 2018.
- [6] S. D. Kamil, D. Widiyanto, dan N. Chamidah, “Perbandingan Metode Decision Tree dengan Naive Bayes dalam Klasifikasi Tumor Otak Citra MRI,” in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2020, hal. 539–550.
- [7] “Peta Risiko,” *Satuan Tugas Penanganan COVID-19*, 2021. <https://covid19.go.id/peta-risiko> (diakses Feb 26, 2021).
- [8] “Data Kasus Covid-19 di Jawa Barat,” *Pusat Informasi dan Koordinasi COVID-19 Provinsi Jawa Barat*, 2021. <https://pikobar.jabarprov.go.id/table-case> (diakses Mar 29, 2021).
- [9] “Update Penanggulangan Covid-19,” *Info Corona Pemerintah Provinsi Bali*, 2021. <https://infocorona.baliprov.go.id/2021/03/28/update-penanggulangan-COVID-19-minggu-28-maret-2021/> (diakses Mar 30, 2021).
- [10] W. Natalia, “Perkembangan Kasus Covid-19 di Kalteng,” *Multimedia Center Provinsi Kalimantan Tengah*, 2021. <https://mmc.kalteng.go.id/berita/read/33542/perkembangan-kasus-COVID-19-di-kalteng-semuh-33-orang-dan-konfirmasi-baru-111-orang> (diakses Mar 31, 2021).
- [11] “Data Harian Covid-19,” https://m.andrafarm.com/_andra.php?_i=daftar-co19-kota&urut=2&asc=00100000000&no1=2&corke=434 (diakses Mar 31, 2021).
- [12] “Peta Sebaran Kasus COVID-19,” *Media Center Satuan Tugas COVID-19 Kalimantan Tengah*, 2021. <https://corona.kalteng.go.id/> (diakses Apr 08, 2021).
- [13] “Tangsel Tanggap COVID-19,” *Pemerintah Kota Tangerang Selatan*, 2021. <https://lawancovid19.tangerangselatankota.go.id/> (diakses Apr 08, 2021).
- [14] “COVID-19 NTB,” 2021. <https://corona.ntbprov.go.id/list-data> (diakses Apr 08, 2021).
- [15] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, 3 ed. Elsevier, 2012.
- [16] M. S. Anggreany, “Confusion Matrix,” *Binus University*. <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> (diakses Apr 01, 2021).