

Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 di Kecamatan Medan Helvetia

Riah Ukur Ginting^{1*}, Mirna Yanti Harefa², Rianto Sitanggang³, Immanuel H.G Manurung⁴

^{1,2,3,4}Program Studi Sistem Informasi, Universitas Sari Mutiara Indonesia, Medan, Indonesia

*Penulis Korespondensi : riahukur@gmail.com

Article Info

Received : 03 Desember 2023

Revised : 16 Desember 2023

Accepted : 28 Desember 2023

Abstract : *The development of Science and Technology currently plays an important role in companies and government agencies in providing useful information for the interests of operations and management. Therefore, many companies or government agencies have information technology departments or information systems. Likewise UPT. The Helvetia District Health Center, namely in processing and classifying the spread of Covid-19, must use an information system in order to support activities in handling the spread of the corona virus, especially in the Medan Helvetia District. In this study, a classification system was built to classify the spread of Covid-19 using the C4.5 Algorithm. The development of this system uses the PHP programming language with a MySQL database. The results obtained in the test by applying the C4.5 Algorithm then obtained an Accuracy of 0.859, so that with this system it is expected that the Covid-19 handling task force at UPT. The Helvetia District Health Center can classify the spread of Covid-19 in the Medan Helvetia District to be more efficient and effective.*

Abstrak : Perkembangan Ilmu Pengetahuan dan Teknologi saat ini sangat memegang peranan penting dalam perusahaan maupun instansi pemerintahan dalam menyediakan informasi yang berguna bagi kepentingan operasi maupun manajemen. Oleh karena itu, banyak perusahaan atau instansi pemerintahan yang memiliki departemen teknologi informasi atau sistem informasi. Begitu juga UPT. Puskesmas Kecamatan Helvetia yaitu dalam mengolah dan mengklasifikasikan penyebaran Covid-19 harus menggunakan sistem informasi agar dapat menunjang kegiatan penanganan penyebaran virus corona khususnya di wilayah Kecamatan Medan Helvetia. Pada penelitian ini dibangun sistem klasifikasi untuk mengklasifikasikan penyebaran Covid-19 dengan menggunakan Algoritma C4.5. Pengembangan sistem ini menggunakan bahasa pemrograman PHP dengan database MySQL. Hasil yang diperoleh dalam pengujian dengan menerapkan Algoritma C4.5 maka diperoleh Accuracy sebesar 0.859, sehingga dengan adanya sistem ini diharapkan satuan tugas penanganan Covid-19 di UPT. Puskesmas Kecamatan Helvetia dapat mengklasifikasikan penyebaran Covid-19 di Kecamatan Medan Helvetia menjadi lebih efisien dan efektif.

Keyword : *Algorithm, Classification of Covid-19 Spread, MySQL, PHP.*

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. Dunia saat ini sedang waspada terhadap penyebaran virus yang dikenal dengan virus corona, Virus Corona (CoV) adalah bagian dari penyakit yang

menyebabkan flu hingga penyakit yang lebih parah seperti MERS-CoV (*Middle East Respiratory Syndrome*), penyakit yang disebabkan oleh virus corona sering dikenal dengan istilah Covid-19 yang ditemukan pada tahun 2019 di Wuhan, China [1]. COVID-19 (*Corona Virus Diseases-19*) merupakan penyakit menular yang disebabkan oleh *corona virus* jenis baru yaitu SARS-CoV-2 dan ditandai dengan adanya gejala seperti demam, batuk dan sesak nafas [2].

Penelitian sebelumnya dengan judul Klasifikasi Penyebaran COVID-19 Menggunakan Algoritma C4.5 Kota Pagar Alam. Penelitian ini membahas tentang analisis pola persebaran COVID-19 dengan menggunakan Algoritma C4.5. Penerapan Algoritma C4.5 pada penelitian ini memiliki tingkat akurasi yang sangat baik dalam mengklasifikasikan penyebaran COVID-19 di Kota Pagar Alam dan menghasilkan informasi yang akurat [7].

Studi Literatur

Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer (machine learning)* untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases (KDD)* adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD [8].

Klasifikasi

Klasifikasi adalah proses membedakan kelas a yang memiliki tujuan agar dapat

digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (13).

Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area *machine learning* yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3. Algoritma C4.5 dan ID3 model yang tak terpisahkan, karena membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5 [15]. 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;

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

Corona Virus Diseases 2019 (COVID 19)

Corona Virus Disease atau 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 *severe acute respiratory 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 terkontaminasi *droplet* tersebut tersentuh oleh orang yang ada disekitar. Pada tanggal 11 Maret 2020 Badan Kesehatan

Dunia atau *World Health Organization* menetapkan COVID-19 ataupun virus corona sebagai *pandemic global*[19].

UML (Unified Modeling Language)

Unified Modeling Language (UML) adalah suatu teknik untuk memodelkan sistem. UML ditemukan oleh Grady Booch, Ivar Jacobson, dan James Rumbaugh. UML versi terbaru, yaitu versi 2.5, terdiri dari lima belas diagram. Diagram-digram tersebut dibagi ke dalam dua kelompok, yaitu *structure diagram* dan *behavior diagram*. *Structure diagram* menggambarkan data dan hubungan statis dalam suatu sistem informasi. *Structure diagram* terdiri dari *package*, *object*, *component*, *class*, *deployment*, *composite structure*, dan *profile diagram*. *Behavior diagram* menggambarkan hubungan dinamis diantara objek yang mewakili sistem informasi bisnis. *Behavior diagram* terdiri dari *sequence*, *timing*, *interaction overview*, *activity*, *use case*, *protocol state machine*, *communication* dan *behavior state machine diagram*[20].

Basis Data (Database)

Basis data merupakan sistem yang terdiri atas kumpulan *file* (tabel) yang saling berhubungan (dalam sebuah basis data di sebuah sistem komputer) dan sekumpulan program (DBMS) yang memungkinkan beberapa pemakai dan/atau program lain untuk mengakses dan memanipulasi *file-file* (tabel-tabel) tersebut [21]

PHP (Hypertext Preprocessor)

Pada awalnya PHP merupakan singkatan dari *Personal Home Page*. Sesuai dengan namanya, PHP digunakan untuk membuat website pribadi. Setelah beberapa tahun perkembangannya, PHP menjelma menjadi bahasa pemrograman web yang *powerfull* dan tidak hanya digunakan untuk membuat halaman web sederhana, tetapi juga website populer yang digunakan oleh jutaan orang, seperti

Wikipedia, *Wordpress*, *Joomla*, dan lain-lain.

MySQL (My Structured Query Language)

DBMS (*Data Base Management System*) yang *open source* dengan dua bentuk lisensi, yaitu *free software* (perangkat lunak bebas) dan *shareware* (perangkat lunak berpemilik yang penggunaannya terbatas. Jadi *MySQL* adalah *database server* yang gratis dengan lisensi *GNU General Public License* (*GPL*) sehingga dapat dipakai untuk keperluan pribadi atau komersial tanpa harus membayar lisensi yang ada [26].

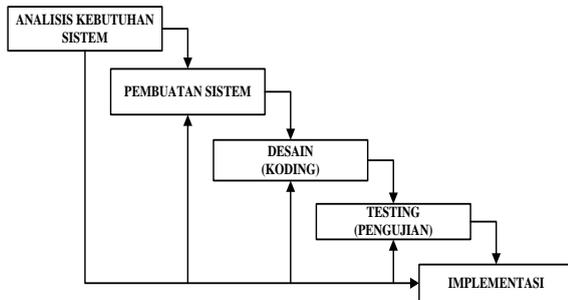
Pengujian Menggunakan Black Box

Black Box Testing merupakan salah satu metode yang mudah digunakan karenahanya memerlukan batas bawah dan batas atas dari data yang di harapkan, estimasi banyaknya data uji dapat dihitung melalui banyaknya field data entri yang akan diuji, aturan entri yang harus dipenuhi serta kasus batas atas dan batas bawah yang memenuhi. Dengan metode ini dapat diketahui jika fungsionalitas masih dapat menerima masukan data yang tidak diharapkan maka menyebabkan data yang disimpan kurang valid[28].

METODE

Penelitian ini dilakukan di UPT. Puskesmas Kecamatan Medan Helvetia yang berada di Jln. Matahari Raya No. 47, Helvetia Tengah, Kecamatan Medan Helvetia, Kota Medan. Pembangunan sistem secara keseluruhan dilakukan melalui beberapa tahapan atau langkah. Metode pengembangan perangkat lunak dikenal juga dengan istilah *software Development Life Cycle* (SDLC). Namun untuk model pengembangan sistem yang digunakan dalam penelitian ini adalah metode *waterfall*. Metode *waterfall* merupakan pendekatan SDLC paling awal yang digunakan untuk pengembangan

perangkat lunak. Urutan dalam metode *waterfall* bersifat serial yang dimulai dari proses perencanaan, analisa, desain, implementasi dan sampai pada implementasi sistem. Berikut ini merupakan gambar model pengembangan sistem dengan metode *waterfall*.



Gambar 1. Model Pengembangan Sistem Metode *Waterfall*

Berdasarkan Gambar 1 diatas maka dapat diuraikan langkah-langkah model pengembangan sistem dengan metode *waterfall* antara lain sebagai berikut:

1. Analisis Kebutuhan Sistem

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya dapat diidentifikasi untuk membangun sebuah aplikasi data mining untuk klasifikasi penyebaran COVID-19 di wilayah Kecamatan Medan Helvetia dengan batasan variabel pada level fungsional yang meliputi data penyebaran penyakit COVID-19 setiap kelurahan, data yang meninggal, data yang terinfeksi, dan data daerah rawan dan tidak rawan.

2. Pembuatan Sistem

Pada tahap ini dilakukan rumusan masalah yang dihadapi oleh Satuan Tugas (SATGAS) COVID-19 di Kecamatan Medan Helvetia dalam penerapan aplikasi data mining klasifikasi penyebaran COVID-19 di wilayah Kecamatan Medan Helvetia menggunakan Algoritma C4.5.

3. Desain (Koding)

Setelah memahami sistem yang ada, maka pada tahap ini dilakukan desain sistem yang akan dibangun seperti UML, perancangan basis data (*database*), perancangan antarmuka, perancangan masukan data dan perancangan keluaran data.

4. Testing (Pengujian)

Pada tahap testing atau pengujian sistem meliputi proses pengujian dengan data riil yang sudah ada. Aplikasi data mining untuk klasifikasi penyebaran COVID-19 yang dibuat apakah sudah bisa digunakan untuk mengklasifikasikan penyebaran COVID-19 di Kecamatan Medan Helvetia. Pengujian dilakukan pada data yang terintegrasi dengan beberapa bagian dengan menggunakan aplikasi *black box*.

5. Implementasi

Pada tahap ini yang dilakukan adalah pembuatan program aplikasi, yang nantinya akan dapat diimplementasikan di Satuan Tugas Penanganan COVID-19 pada Kecamatan Medan Helvetia, Kota Medan untuk mengelola klasifikasi penyebaran COVID-19.

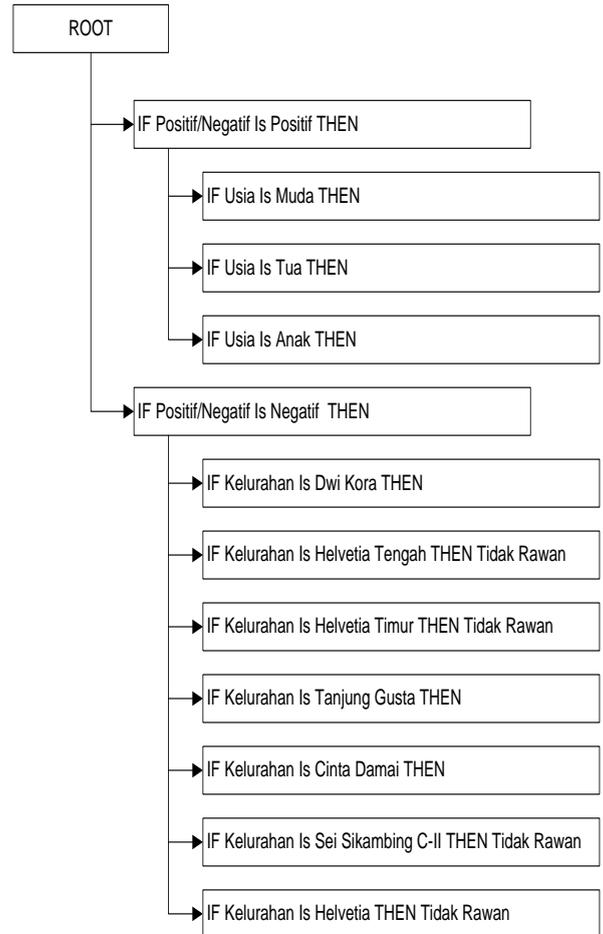
HASIL DAN PEMBAHASAN

Berdasarkan perhitungan nilai *entropy*, *gain*, *split info* dan *gain ratio* diatas, maka dapat ditampilkan dalam bentuk tabel berikut ini:

Tabel 1 Hasil Perhitungan Node 1.2

Node	Nilai Atribut	Total Data	Rawan	Tidak Rawan	Entropy	Gain	Split Info	Gain Ratio
1.2	TOTAL	19	4	15	0.742			
Kelurahan	Cinta Damai	2	1	1	1	0.321	2.65	0.121
	Dwi Kora	4	1	3	0.811			
	Helvetia	1	0	1	0			
	Helvetia Tengah	5	0	5	0			
	Helvetia Timur	2	0	2	0			
	Sei Sikambang C-II	2	0	2	0			
	Tanjung Gusta	3	2	1	0.918			
Jenis Kelamin	Laki-laki	7	2	5	0.863	0.014	0.949	0.015
	Perempuan	12	2	10	0.65			
Usia	Anak	4	1	3	0.811	0.025	1.529	0.016
	Muda	8	1	7	0.544			
	Tua	7	2	5	0.863			
Keadaan Pasien	Sembuh	14	2	12	0.592	0.051	0.831	0.061
	Meninggal	5	2	3	0.971			

Berdasarkan hasil perhitungan *node* 1.2 pada tabel diatas, maka dapat diketahui pohon keputusan untuk *node* 1.2 adalah sebagai berikut:



Gambar 2. Pohon Keputusan Node 1.2

1. Uji Coba *Form* Halaman Utama

Uji coba *form* halaman utama merupakan tampilan awal system aplikasi data mining pengklasifikasian penyebaran Covid-19 di Kecamatan Medan Helvetia.



Gambar 3. Login Covid-19

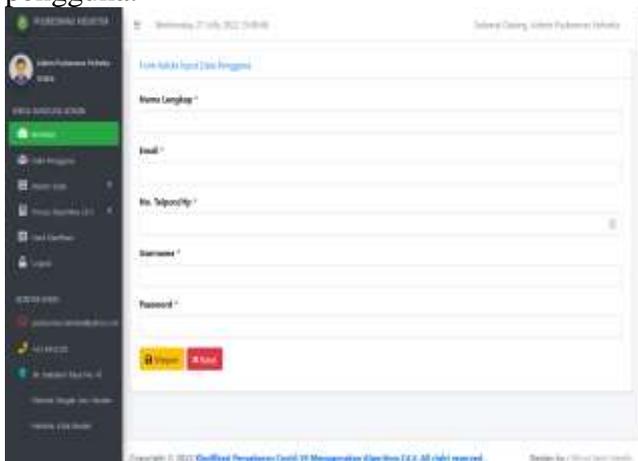
2. Uji Coba *Form* Login

Uji coba *form* login digunakan untuk pengamanan sistem agar sistem selalu terproteksi dari pengguna yang tidak bertanggung jawab dengan cara mengklik tombol “Klik Untuk Login Sistem” di halaman utama.



Gambar 4. Proses Login Uji Coba Form Kelola Data Pengguna

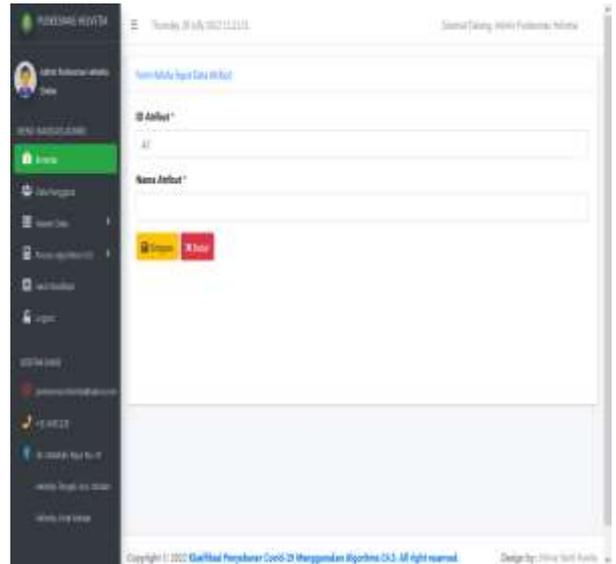
Uji coba ini dilakukan untuk menambah data pengguna dengan tujuan untuk menambahkan data pengguna yang berfungsi untuk akun login pengguna. Kemudian pada uji coba *form* ini dapat ditambah data pengguna baru, edit dan hapus data pengguna.



Gambar 5. Proses Tambah Data Pengguna

Gambar 6 Proses Hapus Data Pengguna Uji Coba *Form* Kelola Data Atribut.

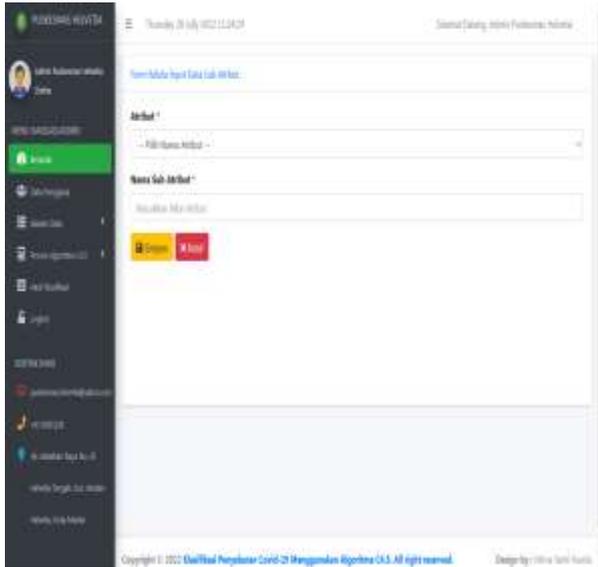
Uji coba ini dilakukan untuk menambah data atribut klasifikasi penyebaran Covid-19 yang baru. Kemudian pada uji coba *form* ini dapat ditambah data atribut, edit dan hapus data atribut.



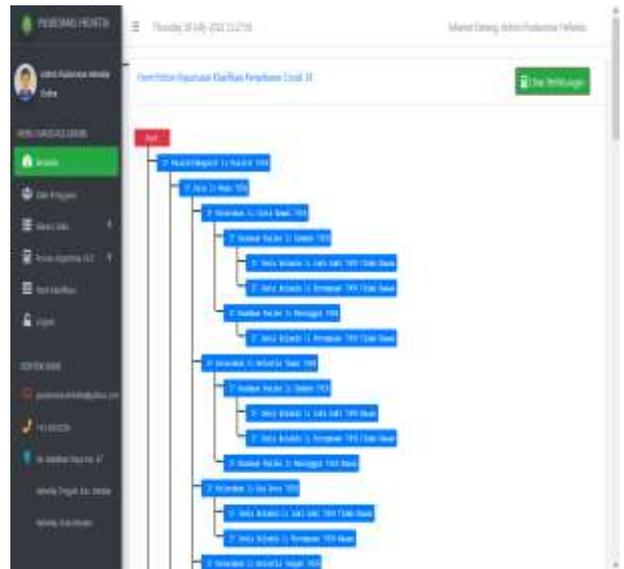
Gambar 6. Proses Hapus Data Pengguna Uji Coba Form Kelola Data Atribut

Uji Coba *Form* Kelola Data Sub Atribut

Form uji coba kelola data kelas berfungsi untuk mengelola data sub atribut. Pada uji coba *form* ini, admin dapat melakukan penambahan data sub atribut, edit dan hapus.

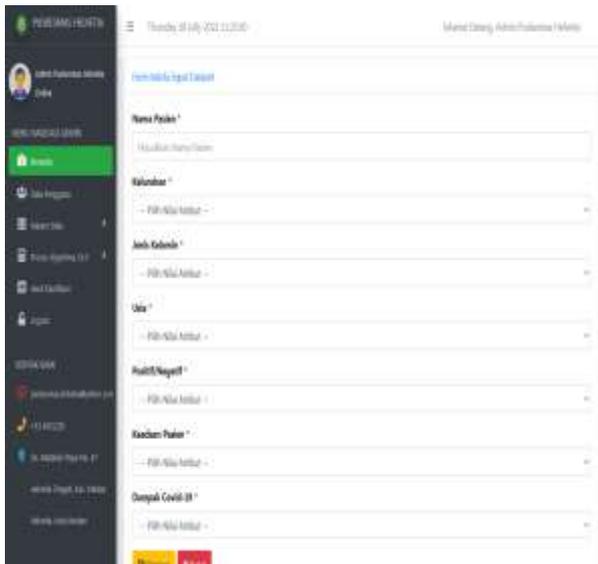


Gambar 7. Proses Tambah Sub Atribut



Gambar 9. Proses Pohon Keputusan

Uji Coba *Form* Kelola Data Covid-19
Uji coba ini dilakukan untuk menambah data penyebaran Covid-19 di wilayah Kecamatan Medan Helvetia.



Gambar 8. Proses Tambah Dat Covid-19

Uji Coba *Form* Kelola Data Pohon Keputusan
Uji coba ini dilakukan untuk menambah data pohon keputusan dengan tujuan untuk menampilkan data pohon keputusan klasifikasi penyebaran Covid-19 menggunakan Algoritma C4.5 pada Kecamatan Medan Helvetia.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada bab sebelumnya dapat disimpulkan sebagai berikut:

1. Sistem informasi pengklasifikasian penyebaran Covid-19 ini dirancang dengan menggunakan pemodelan UML dengan diagram *user case*, *activity* diagram, dan *class* diagram serta sistem ini dibangun dengan menggunakan bahasa pemrograman PHP dengan *database* MySQL.
2. Tahap analisis yang dilakukan adalah dengan menganalisa proses pengklasifikasian penyebaran Covid-19 di wilayah Kecamatan Medan Helvetia.
3. Berdasarkan pengujian sistem terhadap data klasifikasi penyebaran Covid-19 di Kecamatan Medan Helvetia, maka diperoleh pohon keputusan dimana penyebaran Covid-19 hampir seluruh kelurahan berada pada zona tidak rawan.
4. Algoritma C4.5 telah berhasil diterapkan pada klasifikasi penyebaran Covid-19 di Kecamatan Medan Helvetia dengan memperoleh tingkat

akurasi pada *confusion matrix* sebesar 85.9%.

Saran

Berdasarkan kesimpulan diatas, demi lebih optimalnya proses pengklasifikasian penyebaran Covid-19 di Kecamatan Medan Helvetia, maka dapat diberikan saran-saran sebagai berikut:

1. Agar sistem informasi pengklasifikasian penyebaran Covid-19 ini dapat diterapkan dengan baik, dengan dukungan jaringan yang telah dibuat sehingga dapat membantu dalam proses penyampaian informasi klasifikasi penyebaran Covid-19.
2. Diharapkan aplikasi sistem informasi pengklasifikasian Covid-19 ini dapat dikembangkan untuk dapat digunakan oleh satuan tugas penanganan Covid-19 sehingga dapat mengatasi proses penyampaian informasi pengklasifikasian penyebaran Covid-19 dengan mudah, cepat dan efisien.
3. Agar diperoleh hasil yang lebih maksimal lagi dalam mengklasifikasikan penyebaran Covid-19, maka dapat digunakan Algoritma yang lain atau dengan penggabungan kedua Algoritma pada data mining.
4. Pelatihan administrator untuk mempermudah dalam pengelolaan aplikasi data mining ini perlu dilakukan agar proses entri data, pengolahan data dan penyimpanan data dapat berjalan dengan lebih cepat, efisien dan akurat.
Perlu adanya perencanaan ruang khusus untuk ruang kerja bagi admin pengelola aplikasi sistem pengklasifikasian penyebaran Covid-19 ini nantinya, agar sistem berjalan lancar dan efisien.

DAFTAR PUSTAKA

[1] F. Fitriyadi dan M. Muqorobin, "Prediction System for the Spread of Corona Virus in Central Java with K-

Nearest Neighbor (KNN) Method," *IJCIS*, vol. 2, no. 3, hlm. 80–85, Agu 2021, doi: 10.29040/ijcis.v2i3.41.

[2] F. Salsabila dan S. M. Intani, "IMPLEMENTASI ALGORITMA K-MEANS DAN C4.5 DALAM MENENTUKAN TINGKAT PENYEBARAN COVID-19 DI INDONESIA," *Seri Sains dan Teknologi*, vol. 7, no. 1, hlm. 6, 2021.

[3] S. M. Turnip dan P. Silitonga, "Analisis Pola Penyebaran Penyakit dengan Menggunakan Algoritma C4.5," *JTIUST*, vol. 3, no. 1, Jun 2018.

[4] D. Sartika dan Y. Yupianti, "Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus: Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna)," *Rekayasa*, vol. 13, no. 1, hlm. 71–76, Mar 2020, doi: 10.21107/rekayasa.v13i1.5912.

[5] R. Muttaqien, M. G. Pradana, dan A. Pramuntadi, "Implementation of Data Mining Using C4.5 Algorithm for Predicting Customer Loyalty of PT. Pegadaian (Persero) Pati Area Office," *IJCIS*, vol. 2, no. 3, hlm. 64–68, Agu 2021, doi: 10.29040/ijcis.v2i3.36.

[6] N. Anwar, A. Pranolo, dan R. Kurnaiwan, "Grouping the community health center patients based on the disease characteristics using C4.5 decision tree," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 403, hlm. 012084, Okt 2018, doi: 10.1088/1757-899X/403/1/012084.

[7] R. S. Putra, E. D. Putra, M. H. Rifqo, dan H. Witriyono, "Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 Kota Pagar Alam," *JUKOMIKA*, vol. 4, no. 1, hlm. 13, 2021.

[8] F. Riandari dan A. Simangunsong, *Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa*. Kota Medan: CV. Rudang Mayang, 2019.

[9] D. Sepri dan Y. Fimazid, "Pengelompokan Penyebaran Covid-19 di Kota Padang Menggunakan Algoritma K-Medoids," *Insearch*, vol. 1, no. 2, hlm. 7, 2021.

[10] A. A. K. Hartama, "Klasifikasi Penyakit Hipertensi Menggunakan Algoritma C4.5 Studi Kasus RSUD. Provinsi NTB," Universitas Sanata

Dharma, Yogyakarta, 2017. [Daring]. Tersedia pada: https://repository.usd.ac.id/16526/2/125314119_full.pdf.