

Implementasi Algoritma C4.5 untuk Medeteksi Golongan Penderita Covid-19 Berdasarkan Gejala dan Penyebabnya

Efani Desi ^{1*}, Siti Aliyah ²

^{1,2} Universitas Potensi Utama, Kota Medan, Provinsi Sumatera Utara, Indonesia.

article info

Article history:

Received 4 January 2023

Received in revised form

13 March 2023

Accepted 1 May 2023

Available online July 2023

DOI:

<https://doi.org/10.35870/jti.k.v7i3.872>

Keywords:

C4.5 Algorithm; Data Mining; Covid-19; ODP; PDP; OTG.

abstract

Corona virus disease/Covid-19 is the name given by the World Health Organization (WHO) for sufferers infected with the 2019 novel corona virus. Seeing the development of the spread of the virus which continues to increase and the rate of its spread is even faster, it has created anxiety in all circles in public. Therefore, an appropriate method is needed that can make it easier to detect Covid-19 sufferers without having to do Covid-19 RT-PCR. So that the treatment of patients can be handled properly. The author uses the C4.5 Algorithm method which can predict whether the patient has Covid-19 based on the symptoms and causes. The results of this study researchers used the RapidMiner application to apply the C4.5 Algorithm method so that the results obtained showed an accuracy value of 70% for Covid-19 sufferers. So that with this research the public can self-detect and treat the symptoms of Covid-19 suffered if the patient belongs to the ODP group, the action that must be taken is to self-isolate for 14 days to carry out the recommended treatment and to avoid spreading the virus to the surrounding environment, so as well as other patient groups.

abstrak

Kata Kunci:

Algoritma C4.5; Data Mining; Covid-19; ODP; PDP; OTG.

Corona virus disease/Covid-19 merupakan nama yang diberikan Wolrd Health Organization (WHO) untuk penderita yang terinfeksi oleh virus novel corona 2019. Melihat perkembangan penyebaran virus yang terus menerus meningkat dan tingkat penyebarannya pun semakin cepat sehingga membuat keresahan disemua kalangan dimuka publik. Oleh karena itu, diperlukanya suatu metode yang tepat yang dapat mempermudah untuk mendeteksi penderita Covid-19 tanpa harus melakukan RT-PCR Covid-19. Sehingga penanganan terhadap penderita dapat segera ditangani dengan baik. Penulis menggunakan metode Algoritma C4.5 yang dapat memprediksi apakah pasien tersebut menderita Covid-19 berdasarkan gejala dan penyebabnya. Hasil penelitian ini peneliti menggunakan aplikasi RapidMiner untuk menerapkan metode Algoritma C4.5 sehingga hasil yang didapat menunjukkan nilai akurasi penderita Covid-19 sebanyak 70%. Sehingga dengan adanya penelitian ini publik bisa mendeteksi diri serta menangani gejala Covid-19 yang diderita apabila pasien tersebut termasuk golongan ODP maka tindakan yang harus dilakukan adalah melakukan isolasi diri selama 14 hari untuk melalukan pengobatan yang telah dianjurkan serta menghindari penyebaran virus terhadap lingkungan sekitarnya, begitu juga dengan golongan penderita yang lainnya.

Author. Email: efanidesi88@gmail.com ^{1}, aliyahsiti478@gmail.com ².

© E-ISSN: 2580-1643.

Copyright @ 2023. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET)
(<http://creativecommons.org/licenses/by/4.0/>).

1. Latar Belakang

Pada awal tahun 2020, dunia digemparkan dengan kejadian infeksi berat dengan penyebab yang belum diketahui, kejadian ini berasal dari laporan yang diberikan oleh China kepada *World Health Organization* (WHO) dimana laporan tersebut menyatakan terdapat 44 pasien pneumonia yang berat di Kota Wuhan tepatnya di akhir tahun 2019. Dugaan awal terkait dengan infeksi ini adalah pasar basah yang menjual ikan, hewan laut dan berbagai hewan lain. Berdasarkan kejadian infeksi ini menunjukkan hubungan yang dekat dengan virus corona penyebab *Severe Acute Respiratory Syndrome* (SARS) yang mewabah di Hongkong pada tahun 2003 silam, sehingga WHO menamakannya sebagai novel corona virus (nCoV-19). Ancaman virus ini semakin besar ketika terdapat kasus yang menunjukkan bahwa terdapat penularan antar manusia yang terjadi pada dokter dan petugas medis yang merawat pasien tanpa ada riwayat berpegian. Di sisi lain terdapat Laporan lain yang menunjukkan penularan pada pendamping wisatawan China yang berkunjung ke Jepang. Sehingga menimbulkan peningkatan jumlah kasus yang sangat drastis pada Januari 2020 dimana terdapat peningkatan 2000 kasus terkonfirmasi dalam 24 jam. Pada akhir Januari 2020 WHO menyatakan status Global Emergency pada kasus virus Corona dan pada 11 Februari 2020 WHO menamakannya sebagai Covid-19 [1].

Klasifikasi infeksi *Covid-19* di Indonesia pada saat ini didasarkan pada buku panduan tata laksana pneumonia Covid-19 Kementerian Kesehatan Republik Indonesia (Kemenkes RI). Berdasarkan klasifikasinya pasien penderita Covid-19 digolongkan menjadi beberapa golongan berdasarkan gejala yang dialami pasien dan penyebabnya seperti ODP dengan gejala Infeksi Saluran Pernapasan Akut (ISPA) yaitu demam ($\geq 38^{\circ}\text{C}$) atau riwayat penyakit demam yang diikuti dengan gejala gejala penyakit pernapasan seperti: batuk/ sesak nafas/ sakit tenggorokan/ pilek/ pneumonia ringan hingga berat. Dan PDP dengan gangguan sistem pernapasan dan memiliki riwayat kontak dengan kasus konfirmasi COVID-19, serta OTG dimana orang tanpa gejala dengan riwayat kontak erat dengan kasus konfirmasi COVID-19 [1].

Menurut penyebaran dan perkembangan dari Covid-19 ini aktifitas di luar rumah harus dibatasi agar penularannya tidak semakin parah. Sekalipun dalam berbagai pencegahan untuk menanggulangi terjangkitnya Covid-19 dengan demikian kemungkinan yang akan terjadi bagi seseorang terinfeksi Covid-19 sangat rentan terjadi. Oleh karena itu diperlukan suatu metode agar memudahkan bagi masyarakat umum untuk deteksi diri mereka apakah dirinya terinfeksi Covid-19 tanpa harus melakukan test RT-PCR untuk menanggulangi kemungkinan terburuk dari terinfeksi Covid-19. Maka Algoritma C4.5 adalah metode yang bisa dipakai dalam mendeteksi kemungkinan seseorang terjangkit Covid-19 dari gejala dan penyebabnya Dengan demikian, gejala dari penderita Covid-19 segera dapat dideteksi dan ditangani sejak awal munculnya gejala tersebut [2].

Salah satu penyakit yang saat ini paling cepat menular pada abad ke-21 adalah Covid-19. Agar penyakit tersebut terdeteksi dengan cepat digunakan pengujian RT-PCR, CT scan, serta menggunakan gambar *Chest X-Ray* (CXR). Pemindai CT (*Computed Tomography*) dan juga dengan pengujian RT-PCR tidak tersedia di sebagian besar rumah sakit. Oleh karena itu dalam berbagai kasus gambar CXR adalah alat dengan biaya yang terjangkau untuk membantu para dokter dalam mengambil keputusan untuk diagnosa Covid-19. Jaringan saraf tiruan dalam hal ini memungkinkan dipelajari untuk membangun sistem triase Covid-19 terutama untuk pembelajaran dalam memiliki potensi besar untuk pasien yang terinfeksi Covid-19 yang masih tergolong rendah. Rumah Sakit Universitario Clínico San Cecilio, Granada, Spanyol, membangun COVIDGR-1.0, homogen dan database seimbang dengan cakupan semua tingkat keparahan, dari normal dengan RT-PCR Positif, Ringan, Sedang hingga Berat. CovidGR-1.0 berisi 426 tampilan CXR PA (PosteroAnterior) positif dan 426 negatif dan diusulkan metodologi Jaringan Berbasis Data Cerdas Covid (Covid-SDNet) untuk meningkatkan kapasitas generalisasi model klasifikasi Covid. COVIDGR-1.0 beserta label tingkat keparahannya tersedia untuk komunitas ilmiah melalui tautan ini [3].

Di bulan Desember 2019, Covid-19 menyerang kota Wuhan yang berada di Tiongkok dan telah dilaporkan di banyak negara dengan jutaan orang terinfeksi hanya dalam waktu empat bulan. *Tomografi* terkomputasi

dada (CT) telah terbukti menjadi suplemen yang berguna untuk membalikkan reaksi berantai polimerase transkripsi (RT-PCR) sehingga telah terbukti memiliki sensitivitas tinggi untuk mendiagnosis kondisi ini. Pemeriksaan radiologi menjadi salah satu pemeriksaan dini bagi pasien yang infeksi COVID-19. Penemuan CT telah diusulkan sebagai bukti penting untuk pemeriksaan ilmiah COVID-19 di Hubei, China. Akan tetapi pengelompokan pasien dengan gambar CT dada adalah hal yang sulit. Dengan demikian, penelitian ini, jaringan memori jangka pendek dua arah yang dalam dengan model jaringan kepadatan campuran (DBM) diusulkan, *algoritma Memetic Adaptive Differential Evolution* (MADE) digunakan yang mempertimbangkan kumpulan data *tomografi dada-Computed Tomography (chest-CT)*. Analisis komparatif mengungkapkan bahwa model MADE-DBM yang diusulkan mengungguli pendekatan klasifikasi COVID-19 kompetitif dalam hal berbagai metrik kinerja secara *real-time* (Pathak Y. et al., 2021).

Bertambahnya serangan jaringan terhadap Internet, penulis memakai algoritma klasifikasi pohon keputusan C4.5 sebagai deteksi untuk 7 jenis serangan jaringan dari perspektif lalu lintas jaringan. Yang diproses pertamakali adalah data, selanjutnya mengekstrak 62 fitur dan terakhir memakai algoritma C4.5 yang berfungsi untuk mendeteksi lalu lintas. Eksperimen ini menggunakan kumpulan data publik CSE-CIC-IDS2018 untuk verifikasi. Hasil percobaan menunjukkan bahwa metode dalam artikel ini dapat secara efektif mendeteksi berbagai jenis serangan dunia maya. Tingkat akurasi bisa mencapai 96,7%, dan tingkat positif palsu hanya 4,5% (Zhou J. et al., 2021).

Covid-19 adalah virus varian baru yang muncul di kota Wuhan pada tahun 2019. Virus ini menyebar dengan cepatnya dan tidak terkendali bahkan hingga sampai ke Indonesia. Dalam upaya menaggulangi virus COVID-19 maka dilakukan test PCR dan antigen untuk deteksi virus tersebut. Dalam penelitian ini menggunakan metode Naïve Bayes, C4.5 dan K-Nearest Neighbor dimana tingkat akurasinya tinggi. Hasil yang didapatkan dari penelitian ini yaitu dimana algoritma C4.5 dalam mengolah data sekunder dari Dinas Kesehatan Kota Dumai dalam memprediksikan satatus dari

pasien yang terinfeksi COVID-19 dengan hasil akurasi 86,54%, recall 71,51% dan presisi 85,14%. Penelitian ini berimplikasi pada peneliti selanjutnya dalam memilih algoritma untuk memprediksi kasus COVID-19 [4].

Peran orang tua dalam pelaksanaan sekolah dirumah selama pandemi COVID-19 sangatlah penting bagi anak yang masih TK dan SD. Karena aktivitas, dukungan, penagmabilan keputusan dan nasihat dalam keteladanan di keluarga menjadi faktor penting dalam tumbuh kembang anak. Demografi adalah salah satu faktor berpengaruh dalam keterlibatan orang tua. Penelitian ini digunakan untuk mengelompokkan tingkat keterlibatan orang tua berlandaskan faktor demografi yang menggunakan algoritma C4.5. Diawali dari menyusun skala keterlibatan orang tua kemudian mengelompokkan skor keterlibatan orang tua yang sangat tinggi, tinggi, sedang rendah serta sangat rendah. Faktor demografi setelah dilakukan *preprocessing* data digunakan sebagai input pada algoritma C4.5. Pohon yang dikembangkan menunjukkan bahwa akar pohon tersebut merupakan tingkat pendidikan ibu. Hasil pengujian menunjukkan akurasi algoritma C4.5 dalam mengklasifikasikan keterlibatan orang tua sebesar 86,67%. Hasil ini cukup baik untuk digunakan dalam proses klasifikasi [5].

Organisasi Kesehatan Dunia (WHO) menyatakan wabah penyakit *Coronavirus 2019 (Covid-19)* sebagai pandemi dan menjadikannya trending topik di *platform* media sosial, seperti Facebook dan Twitter. Oleh karena itu, karya ini bertujuan untuk mengusulkan sebuah model pembelajaran *ensemble* yang dapat mengklasifikasikan sentimen di dalam opini masyarakat dari Twitter. Model *ansambel* menggunakan *Naïve Bayes Classifier*, C4.5, dan *k-Nearest Neighbours* sebagai pembelajar dasar dengan mekanisme voting untuk menghasilkan keputusan akhir. Untuk pembelajaran model *ansambel* menggunakan data set berisi 3884 data bersih yang berhasil diunduh menggunakan Twitter API terkait pencegahan wabah Covid-19 dan diolah menggunakan metode TF-IDF. Data set memiliki dua kelas, yaitu 'positif' dan 'negatif' untuk mewakili sentimen opini di setiap data. Model yang diusulkan mendapatkan akurasi 80,61%, recall 79,49%, dan presisi 81,20%, setelah dievaluasi menggunakan 10-fold Cross Validation. Ini juga tampil lebih baik jika

dibandingkan dengan beberapa model pembelajaran yang hanya menggunakan algoritma Machine Learning tunggal [6].

COVID-19 dapat menyebar melalui udara, yang disebabkan oleh menghirup tetesan kecil yang mengandung SARS-CoV-2 di lingkungan dalam ruangan. Secara khusus, baik orang dengan gejala maupun orang tanpa gejala membuat banyak tetesan kecil dan tetesan pernapasan saat bernapas, bersin, batuk, atau berbicara. Saat tetesan kecil ini terpapar ke udara di sekitarnya, mereka dapat bereaksi dengan partikel (PM) dan bertahan di udara untuk waktu yang lama. Masa kelangsungan hidup tergantung pada berbagai kondisi, termasuk jenis permukaan, suhu, dan kelembaban relatif. Sistem ventilasi adalah salah satu cara untuk mengatasi masalah ini. Penelitian ini memberikan kontribusi yang signifikan terhadap pengembangan sistem ventilasi cerdas yang memanfaatkan Internet of Things (IoT) dan algoritme pohon keputusan. Fungsi dari sistem ini adalah untuk mendapatkan lingkungan ruangan yang bersih dari virus dengan menyemprotkan disinfektan secara otomatis dan mengontrol suhu dan kelembaban udara. Berdasarkan hasil penelitian ini, sistem mampu mengontrol dan mengkategorikan kondisi ruangan berdasarkan faktor suhu dan kelembaban. Nilai klasifikasi menggunakan algoritma pohon keputusan C4.5 adalah 92,33% dengan rata-rata nilai suhu dan kelembaban 25°C dan 49% [7].

Berdasarkan klasifikasinya pasien penderita COVID-19 digolongkan menjadi beberapa golongan berdasarkan gejala yang dialami pasien dan penyebabnya seperti ODP, PDP dan OTG. ODP atau Orang yang menderita Infeksi Saluran Pernapasan Akut (ISPA) yaitu demam diatas 38°C atau memiliki riwayat penyakit demam yang diikuti dengan gejala penyakit pernapasan seperti: batuk/ sesak nafas/ sakit tenggorokan/ pilek/ pneumonia ringan atau berat dan tidak ada penyebab gambaran klinis yang meyakinkan. PDP adalah orang yang menderita demam dengan suhu diatas 38°C atau memiliki riwayat penyakit demam beserta dengan gejala gangguan pernapasan berupa pilek/sakit tenggorokan/batuk dan tidak ada penyebab berdasarkan gambaran klinis yang meyakinkan dan pada 14 hari terakhir sebelum timbul gejala memiliki riwayat perjalanan atau tinggal di negara/wilayah

yang termasuk tempat yang terinfeksi oleh virus. Sedangkan OTG merupakan orang yang tidak terdapat gejala dan terkonfirmasi menderita COVID-19. Orang tanpa gejala merupakan seseorang dengan riwayat kontak erat dengan kasus konfirmasi COVID-19 [1].

Klasifikasi pohon keputusan adalah pembelajaran berbasis contoh dan metode penambangan data yang efektif, yang diterapkan untuk memproses data kontinu yang besar untuk menghemat waktu pemrosesan. Dalam makalah ini, di satu sisi, metode pemisahan K-equal disajikan, yang mengurangi jumlah perhitungan dan mempercepat kecepatan pembangunan pohon. Di sisi lain, rumus perhitungan rasio perolehan informasi pada algoritma C4.5 disederhanakan. Algoritme C4.5 yang ditingkatkan diterapkan untuk menganalisis perilaku pembelian pengguna, termasuk membantu merekomendasikan produk yang mereka minati, dan mempercepat mereka untuk menemukan barang yang mereka inginkan dengan cepat. Hasil percobaan menunjukkan bahwa metode lanjutan dapat secara efektif meningkatkan kecepatan pelatihan model [8]-[9].

2. Metode Penelitian

Penelitian ini menggunakan algoritma klasifikasi pohon keputusan dianalisis, termasuk algoritme ID3 dan C4.5, diikuti oleh beberapa algoritme yang disempurnakan. Deret Taylor dan deret Maclaurin diterapkan untuk menyederhanakan formula rasio perolehan informasi sehingga efisiensi waktu perhitungan dapat ditingkatkan. Algoritma C4.5 yang ditingkatkan dapat diterapkan dalam sistem rekomendasi pekerjaan siswa. Melalui algoritme pohon keputusan C4.5 yang ditingkatkan ini, lulusan dapat memiliki panduan yang baik saat mencari pekerjaan dan guru juga dapat membantu siswanya dalam pengambilan keputusan dan rekomendasi pekerjaan [9].

Data mining merupakan suatu proses guna menemukan hubungan, pola dan hubungan baru yang bermakna dengan menyaring data yang besar dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika [13]. Definisi lain data mining merupakan rantaian proses yang menjalankan satu

atau lebih teknik pembelajaran komputer guna menganalisis dan mengekstrak pengetahuan secara otomatis [10]. Algoritma C4.5 tanpa metode diskritisasi menghasilkan skor kebaikan 0,2 sedangkan algoritma C4.5 modifikasi menggunakan teknik diskritisasi memperoleh skor kebaikan 2,49 yang lebih tinggi dari yang terakhir. Dengan menggunakan metode diskritisasi, skor kebaikan meningkat membuat algoritma C4.5 efisien dalam menangani variabel kontinu dengan data yang lebih besar [11]. Algoritma C4.5 adalah algoritma lanjutan dari algoritma ID3. Oleh karena Algoritma C4.5 memiliki prinsip dasar kerja yang sama dengan algoritma ID3. Tetapi dalam Algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan Gain Ratio dengan persamaan [12]:

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

Keterangan:

- S : Himpunan kasus,
- A : Atribut,
- N : Jumlah partisi dalam atribut,
- S_i : Jumlah kasus pada partisi ke-I,
- S : Jumlah kasus.

Sedangkan untuk perhitungan nilai Entropy dapat dilakukan dengan rumus:

$$Entropy(S) = \sum_{i=0}^n -P_i \times \log_2 P_i$$

Decision Tree (Pohon Keputusan) merupakan struktur representasi pengetahuan yang terdiri dari simpul-simpul dan cabang yang diorganisasikan dalam bentuk pohon dimana setiap simpul non-daun internal diberi label dengan nilai atributnya. Cabang yang keluar dari internal node diberi keterangan dengan nilai atribut pada node tersebut. Setiap simpul diberi label dengan kelas (nilai tujuan atribut). Model berbasis pohon yang meliputi klasifikasi dan regresi pohon yang merupakan implementasi umum dari pemodelan induksi. Model pohon keputusan merupakan metode yang cocok untuk data mining. *Decision Tree* mudah untuk membangun, mudah ditafsirkan, mudah diintegrasikan dengan sistem database dan mereka memiliki akurasi yang sebanding atau lebih baik aplikasi [2].

3. Hasil dan Pembahasan

Penentuan Status Penderita Covid-19 Berdasarkan Gejala dan Penyebab

Di dalam penelitian ini terdapat 20 pasien yang memiliki gejala serta penyebab yang diduga merupakan penderita COVID-19. Dimana penderita akan digolongkan menjadi 3 bagian yaitu ODP, PDP, dan OTG dengan tabel 1 berikut ini :

Tabel 1. Status Penderita Covid-19 Berdasarkan Gejala Dan Penyebab

Kode	Keterangan
ODP	Demam / Gangguan Pernapasan (Gejala Umum)
PDP	Demam dan Gangguan Pernapasan serta Interaksi Dengan Penyebab Covid-19 Lainnya
OTG	Tanpa Gejala Tetapi Positif Covid-19

Berdasarkan tabel 1 diatas, penderita memiliki gejala dan penyebab yang berbeda-beda sehingga dipenelitian ini, gejala serta penyebabnya pecahkan menjadi beberapa gejala dan penyebabnya agar lebih mudah dalam perhitungan algoritma C4.5 untuk mendeteksi pasien sehingga diperkirakan tergolong dalam golongan penderita ODP, PDP atau OTG yang dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Gejala dan Penyebabnya

Kode	Keterangan
G01	Batuk Kering
G02	Pilek / Flu
G03	Sesak Napas
G04	Demam
G05	Sakit Tenggorokan
P01	Interaksi dengan Penderita COVID-19
P02	Berwisata Ke Tempat Penyebaran COVID-19
P03	Udara (Airborne) / Droplet
P04	Interaksi Dengan Benda Yang Terkontaminasi COVID-19
P05	Limbah Manusia

Berdasarkan kedua tabel datanya maka dapat disimpulkan relasi antara golongan pasien dengan gejala dan penyebabnya sehingga menghasilkan sebuah keputusan status COVID-19 yang dapat dilihat pada tabel 3 berikut ini :

Tabel 3. Keputusan Status COVID-19

Kode	Status Pasien		
	ODP	PDP	OTG
G01	Ya	Ya	Tidak
G02	Ya	Ya	Tidak
G03	Ya	Ya	Tidak
G04	Ya	Ya	Tidak
G05	Ya	Ya	Tidak
P01	Tidak	Ya	Ya
P02	Tidak	Ya	Ya
P03	Tidak	Ya	Ya
P04	Tidak	Ya	Ya
P05	Tidak	Ya	Ya

Dari relasi yang telah didapatkan maka penelitian ini akan menerapkan metode klasifikasi data mining Algoritma C4.5 dan Decision Tree terhadap data pasien dimana data akan diolah dengan menggunakan algoritma C4.5 yang akan menghasilkan model, sehingga model yang dihasilkan tersebut akan dilakukan pengujian menggunakan validation pada aplikasi *RapidMiner* untuk mengetahui nilai akurasi. Maka akan didapat kemungkinan status pasien seperti pada tabel 4 berikut ini :

Tabel 4. Kemungkinan Status Pasien

NO.	G01	G02	G03	G04	G05	P01	P02	P03	P04	P05	Status Pasien
1	Ya	Ya	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
2	Tidak	Ya	Ya	Ya	Ya	Tidak	Ya	Tidak	Tidak	Tidak	PDP
3	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Ya	Tidak	OTG
4	Ya	Ya	Ya	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	PDP
5	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	OTG
6	Tidak	Tidak	Ya	Ya	Tidak	Ya	Ya	Tidak	Tidak	Tidak	PDP
7	Ya	Ya	Ya	Tidak	ODP						
8	Tidak	Ya	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
9	Ya	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
10	Ya	Ya	Ya	Ya	Tidak	Tidak	Tidak	Ya	Ya	Tidak	PDP
11	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	Ya	PDP
12	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Ya	OTG
13	Tidak	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
14	Tidak	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
15	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	OTG
16	Tidak	Ya	Tidak	OTG							
17	Tidak	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	PDP
18	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	OTG
19	Ya	Ya	Ya	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	ODP
20	Ya	Ya	Ya	Tidak	ODP						

Menghitung nilai entropy dan gain dari masing-masing gejala dan penyebabnya

Dari tabel kemungkinan status pasien tersebut, akan diperhitungkan nilai entropy dan gain dari masing-masing gejala dan penyebabnya. Berikut perhitungan nilai entropy dan nilai gain nodenya :

$$Entropy (S) = \sum_{i=0}^n - P_i \times \log_2 P_i.$$

$$Entropy (Total) = \sum_{i=0}^n - P_i \times \log_2 P_i.$$

$$\left(-\frac{8}{20} x \log_2 \left(\frac{8}{20} \right) \right) + \left(-\frac{6}{20} x \log_2 \left(\frac{6}{20} \right) \right) + \left(-\frac{6}{20} x \log_2 \left(\frac{6}{20} \right) \right) = 1,570950594$$

$$Entropy (G1, Ya) = \sum_{i=0}^n - P_i \times \log_2 P_i.$$

$$\left(-\frac{5}{20} x \log_2 \left(\frac{5}{20} \right) \right) + \left(-\frac{2}{20} x \log_2 \left(\frac{2}{20} \right) \right) + \left(-\frac{0}{20} x \log_2 \left(\frac{0}{20} \right) \right) = 0$$

$$\begin{aligned}
 Entropy (G1, Tidak) &= \sum_{i=0}^n -P_i \times \log 2 P_i \\
 &= \left(-\frac{3}{20} \times \log 2 \left(\frac{3}{20} \right) \right) + \left(-\frac{4}{20} \times \log 2 \left(\frac{4}{20} \right) \right) + \\
 &\quad \left(-\frac{6}{20} \times \log 2 \left(\frac{6}{20} \right) \right) = 1,52623491
 \end{aligned}$$

Setelah menghitung entropy maka selanjutnya menghitung gain-nya:

$$Gain (S, A) = Entropy (S) - \sum_{i=0}^n \frac{|S_i|}{|S|} \times Entropy (S_i)$$

$$Gain (Total, G01) = Entropy (Total) - \sum_{i=0}^n \frac{|G01|}{|total|} \times$$

$$\begin{aligned}
 Entropy (G01) &= 1,570950594 - \left(\left(\frac{7}{20} \times 0 \right) + \left(\frac{14}{20} \times 1,52623491 \right) \right) \\
 &= 0,578897903
 \end{aligned}$$

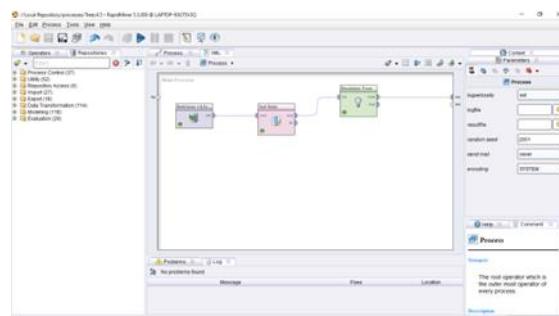
Berdasarkan hasil perhitungan entropy dan gain diatas, maka kita dapat menganalisa perhitungan ketiga jenis gejala tersebut yang dapat dilihat pada tabel 5. Berikut ini :

Tabel 5. Analisa Perhitungan Gejala

Kode	Jlh kasus	ODP	PDP	OTG	Entropy	Gain
G01	Total	20	8	6	1,570950594	0,578897903
	Ya	7	5	2	0	
G02	Tidak	13	3	4	1,52623491	0,78149633
	Ya	9	6	3	0	
G03	Tidak	11	2	3	1,435371391	1,570950594
	Ya	13	8	5	0	
G04	Tidak	7	0	1	6	0,923219672
	Ya	10	5	5	0	
G05	Tidak	10	3	1	6	0,396753422
	Ya	5	3	2	0	
P01	Tidak	15	5	4	1,56559623	0,511615928
	Ya	5	0	3	2	
P02	Tidak	15	8	3	3	0,581290899
	Ya	6	0	3	3	
P03	Tidak	14	8	3	3	0,178982141
	Ya	2	0	1	1	
P04	Tidak	18	8	5	5	0,277089203
	Ya	3	0	1	2	
P05	Tidak	17	8	5	4	0,178982141
	Ya	2	0	1	1	
	Tidak	18	8	5	5	1,546631615

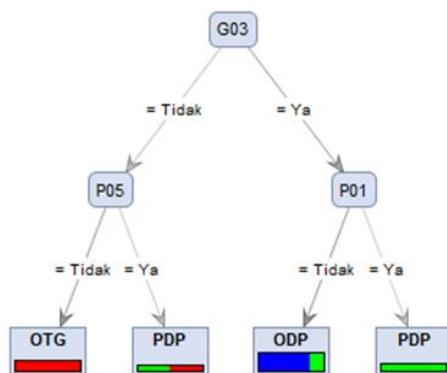
Melakukan uji coba hasil perhitungan entropy dan gain dengan menggunakan aplikasi RapidMiner

Setelah mendapatkan hasil perhitungan nilai entropy dan nilai gainnya, maka akan diuji melalui aplikasi RapidMiner seperti gambar 1 sebagai berikut ini :



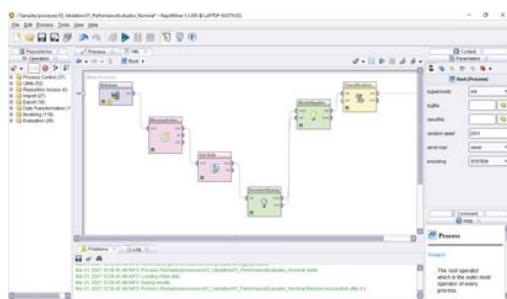
Gambar 1. Hasil uji coba nilai entropy dan gain

Dari hasil uji coba di atas akan terbentuk pohon keputusan seperti gambar 2 berikut ini:



Gambar 2. Decision Tree dari hasil uji coba

Model yang telah dibentuk diuji tingkat akurasinya dengan memasukan data training dengan menggunakan validation pada aplikasi RapidMiner untuk menguji tingkat akurasi. Proses menentukan nilai akurasi tersebut dapat dilihat pada gambar 3 sebagai berikut ini:



Gambar 3. Proses Menentukan Nilai Akurasi pada RapidMiner

Dengan demikian, nilai akurasi yang dihasilkan dengan aplikasi RapidMiner untuk nilai akurasi, nilai *class recall* dan nilai *class precision* yang akan dijelaskan pada gambar 4 sebagai berikut ini:

accuracy:70.00%				
	true ODP	true PDP	true OTG	class precision
pred ODP	8	5	0	61.54%
pred PDP	0	0	0	0.00%
pred OTG	0	1	6	85.71%
class recall	100.00%	0.00%	100.00%	

Gambar 4. Tabel nilai akurasi keseluruhan

Dari gambar 4 di atas dapat dijelaskan bahwa penderita ODP menghasilkan *class recall* sebesar 100% dan *class precision* sebesar 61.54%, sedangkan PDP menghasilkan *class recall* sebesar 00.00% dan *class precision* sebesar 00.00% serta OTG menghasilkan *class recall* sebesar 100.00% dan *class precision* sebesar 85.71% dan nilai akurasi dari perhitungan Algoritma C4.5 tersebut diatas adalah sebesar 70%.

4. Kesimpulan

Dari penelitian ini dapat digaris bawik bahwa mendeteksi golongan penderita COVID-19 dapat dilakukan dengan menerapkan Algoritma C4.5 dimana setelah melakukan uji coba dengan aplikasi RapidMiner menghasilkan tingkat Akurasi sebesar 70 %. Sehingga dengan adanya penelitian ini publik bisa mendeteksi diri serta menangani gejala COVID-19 yang diderita misal apabila dia termasuk golongan ODP maka tindakan yang harus dilakukan adalah melakukan isolasi diri selama 14 untuk melakukan pengobatan yang telah dianjurkan serta menghindari penyebaran terhadap sekitarnya, begitu juga dengan golongan penderita lainnya.

5. Daftar Pustaka

- [1] Handayani, D., Hadi, D.R., Isbaniah, F., Burhan, E. and Agustin, H., 2020. Penyakit virus corona 2019. *Jurnal respirologi indonesia*, 40(2), pp.119-129.
- [2] Sugara, B., Widyatmoko, D., Prakoso, B.S. and Saputro, D.M., 2018. Penerapan Algoritma C4.5 untuk Deteksi Diri Autisme Pada Anak. *Semin. Nas. Teknol. Inf. dan Komun*, 2018, pp.87-96.

- [3] Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J.L., Sevillano-García, I., Rey-Area, M., Charte, D., Guirado, E., Suárez, J.L., Luengo, J., Valero-González, M.A. and García-Villanova, P., 2020. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE journal of biomedical and health informatics*, 24(12), pp.3595-3605. DOI: <https://doi.org/10.1109/JBHI.2020.3037127>.
- [4] Gurning, U.R., Rizaldi, S.T. and Syukron, H., 2022, July. Comparison of Naïve Bayes, C4. 5 and K-Nearest Neighbor for Covid-19 Data Classification. In *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)* (pp. 16-21). IEEE. DOI: <https://doi.org/10.1109/ISITDI55734.2022.9944507>.
- [5] Zaeni, I.A., Anzani, D.R.A., Putra, D.S., Devi, M., Hidayati, L. and Sudjono, I., 2020, September. Classifying the parental involvement on school from home during covid-19 using c4. 5 algorithm. In *2020 4th International Conference on Vocational Education and Training (ICOVET)* (pp. 253-257). IEEE. DOI: <https://doi.org/10.1109/ICOVET50258.2020.9230214>.
- [6] Sancoko, S.D., Diwandari, S. and Fachrie, M., 2022, November. Ensemble Learning for Sentiment Analysis on Twitter Data Related to Covid-19 Preventions. In *2022 International Conference on Information Technology Research and Innovation (ICITRI)* (pp. 89-94). IEEE. DOI: <https://doi.org/10.1109/ICITRI56423.2022.9970234>.
- [7] Febriantono, M.A., Prasetya, N.W. and Husna, N., 2022, October. Intelligent Ventilation System for COVID-19 Prevention using the IoT and Decision Tree. In *2022 8th International Conference on Education and Technology (ICET)* (pp. 91-94). IEEE. DOI: <https://doi.org/10.1109/ICET56879.2022.9990627>.
- [8] Chen, P., 2021, June. Application of an Improved C4. 5 Algorithm in Shopping Websites. In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 63-66). IEEE. DOI: <https://doi.org/10.1109/ICAICA52286.2021.9498086>
- [9] Chen, P., 2021, July. The Application of an Improved C4. 5 Decision Tree. In *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)* (pp. 392-396). IEEE. DOI: <https://doi.org/10.1109/ICNISC54316.2021.00078>
- [10] Azwanti, N., 2018. Analisa Algoritma C4. 5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Inform. Mulawarman J. Ilm. Ilmu Komput.*, 13(1), p.33.
- [11] Mapa, J.S., Sison, A. and Medina, R.P., 2019, December. A Modified C4. 5 Classification Algorithm: With the Discretization Method in Calculating the Goodness Score Equivalent. In *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-4). IEEE. DOI: <https://doi.org/10.1109/ICETAS48360.2019.9117309>.
- [12] Oktavia, C.A., Rahmadwati, R. and Santoso, P.B., 2015. Analisis Kinerja Algoritma C4. 5 Pada Sistem Pendukung Keputusan Penentuan Jenis Pelatihan. *EECCIS*, 9(2), p.60419.
- [13] Pambudi, R.H., Setiawan, B.D. and Indriati, I., 2018. Penerapan Algoritma C4. 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(7), pp.2637-2643.

- [14] Nasrullah, A.H., 2018. Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2), pp.244-250. DOI: <https://doi.org/10.33096/ilkom.v10i2.300.244-250>.
- [15] Rofiuddin, M., 2022. Dampak Corona Virus Disease 19 dan Obligasi Terhadap Nilai Tukar dan Sukuk di Indonesia. *Jurnal Ilmiah Ekonomi Islam*, 8(1), pp.1-9. DOI: <http://dx.doi.org/10.29040/jiei.v8i1.4427>.
- [16] Aditia, A., 2021. Covid-19: Epidemiologi, Virologi, Penularan, Gejala Klinis, Diagnosa, Tatalaksana, Faktor Risiko dan Pencegahan. *Jurnal Penelitian Perawat Profesional*, 3(4), pp.653-660.
- [17] Santoso, A.M.H., 2022. Covid-19: Varian Dan Mutasi. *Jurnal Medika Hutama*, 3(02 Januari), pp.1980-1986.
- [18] Aygün, I., Kaya, B. and Kaya, M., 2021. Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5), pp.2360-2369. DOI: <https://doi.org/10.1109/JBHI.2021.3133103>.
- [19] Pathak, Y., Shukla, P.K. and Arya, K.V., 2020. Deep bidirectional classification model for COVID-19 disease infected patients. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(4), pp.1234-1241. DOI: <https://doi.org/10.1109/TCBB.2020.3009859>.
- [20] Zhou, J., Jiang, X., Liu, C., Zhang, J., Liao, L. and Lu, J., 2021, December. Multi-Traffic Features Network Intrusion Detection Algorithm Based on C4. 5. In *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)* (pp. 548-552). IEEE. DOI: <https://doi.org/10.1109/ICCWAMTIP53232.2021.9674129>.
- [21] Ma, C., Zhang, J., Wang, L. and You, H., 2020, October. Network Attack Detection Based on Kernel Principal Component Analysis and Decision Tree. In *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)* (pp. 84-91). IEEE. DOI: <https://doi.org/10.1109/CyberC49757.2020.00023>.