ANTIVIRUS: Jurnal Ilmiah Teknik Informatika (p – ISSN: 1978 – 5232; e – ISSN: 2527 – 337X)
Vol. 20 No. 1 Mei 2026, pp. 48 – 59

KOMPARASI METODE ALGORITMA C4.5 DAN NAÏVE BAYES TERHADAP KLASIFIKASI STATUS GIZI BALITA DI DESA SUNGAI AMBANGAH KALIMANTAN BARAT

Diterima Redaksi:16 Oktober 2025; Revisi Akhir: 8 Februari 2026; Diterbitkan Online: 31 Mei 2026

Alfina Putri Anggraini[1)], Surti Ayu[2)], Suharyanto[3)], Irmawati[4)]
[1, 2, 3,4)] Universitas Bina Sarana Informatika, Fakultas Teknik dan Informatika
[1, 2,3,4)] Jalan Kramat Raya No.98, Senen. Jakarta Pusat 10450
e-mail: alfinaputrianggraini02@gmail.com[1)], surtiiayu@gmail.com[2)], suharyanto@bsi.ac.id[3)], irmawati.iat@bsi.ac.id[4)]

***Abstrak*: :** *Isu masalah gizi pada anak balita masih menimbulkan tantangan kesehatan yang substansial, khususnya di Desa Sungai Ambangah, Provinsi Kalimantan Barat. Ketidakharmonisan gizi, entah berupa defisiensi atau surplus, mampu memberikan pengaruh buruk terhadap perkembangan tubuh, kemajuan intelektual, serta standar kehidupan anak di kemudian hari. Kajian ini dimaksudkan untuk mengevaluasi efektivitas algoritma C4.5 dan Naive Bayes dalam mengategorikan kondisi gizi balita dengan memanfaatkan dataset dari Pos Kesehatan Desa (Poskesdes) Sungai Ambangah. Pendekatan riset menerapkan metode kuantitatif perbandingan dengan jumlah 1.502 observasi yang telah menjalani pra-pengolahan, meliputi sanitasi data, reduksi data, serta konversi data. Penilaian dilakukan melalui aplikasi RapidMiner dengan prosedur validasi silang 10-fold. Temuan pengujian mengindikasikan bahwa algoritma C4.5 mencapai tingkat ketepatan 96,01% beserta nilai AUC 0,972, sedangkan algoritma Naive Bayes menghasilkan ketepatan 95,54% dengan AUC 0,983. Melalui uji t, nilai p sebesar 0,479 (melebihi 0,05) menyiratkan absennya perbedaan bermakna di antara keduanya. Pada intinya, kedua pendekatan menunjukkan performa yang seimbang dan layak diterapkan sebagai fondasi untuk merancang sistem bantuan pengambilan keputusan dalam penilaian status gizi balita.*

Kata Kunci*—Algoritma C4.5, Data Mining, Klasifikasi, Naïve Bayes, RapidMiner, Status Gizi Balita*

***Abstract*:** *Nutrition problems among young children continue to pose a significant health concern, especially in Sungai Ambangah Village located in West Kalimantan Province. Whether due to deficiencies or overabundances, these nutritional imbalances can negatively impact physical development, cognitive progress, and long-term well-being. The objective of this research is to evaluate the performance of the C4.5 and Naive Bayes algorithms in categorizing the nutritional condition of toddlers, utilizing data collected from the Village Health Post (Poskesdes) in Sungai Ambangah. A quantitative comparative method was applied, involving 1,502 preprocessed observations that included data cleaning, dimensionality reduction, and transformation. Assessment was performed with RapidMiner software employing 10-fold cross-validation. The findings show that the C4.5 algorithm attained an accuracy rate of 96.01% and an AUC of 0.972, whereas the Naive Bayes algorithm reached 95.54% accuracy with an AUC of 0.983. A t-test revealed a p-value of 0.479, which exceeds 0.05, indicating no statistically significant disparity between the algorithms. In summary, both approaches exhibit similar effectiveness and are appropriate for creating a decision support system to evaluate toddlers' nutritional health.*

***Keywords*—** *C4.5 Algorithm, Data Mining, Classification, Naïve Bayes, RapidMiner, Toddler Nutritional Status*

Penulis Korespondensi:
Alfina Putri Anggraini
Program Studi Sistem Informasi
Universitas Bina Sarana Informatika
Email: alfinaputrianggraini02@gmail.com
Orchid ID: https://orcid.org/0009-0001-8421-4238

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

I. PENDAHULUAN

PERMASALAHAN gizi pada balita (bawah lima tahun) masih menjadi isu kesehatan yang serius di berbagai negara, termasuk Indonesia. Berdasarkan hasil Survei Kesehatan Indonesia (SKI) tahun 2024 dari Badan Kebijakan Pembangunan Kesehatan Kemenkes RI yang dapatkan pada situs web badankebijakan.kemkes.go.id, bahwa status gizi dapat dianalisis menggunakan metode *antropometri* untuk menilai kondisi gizi seseorang melalui tiga indikator, yaitu berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB) [1].

Dari hasil survei tersebut, prevalensi status gizi balita di Provinsi Kalimantan Barat, masih ditemukan permasalahan yang cukup serius terkait kondisi gizi anak usia 0 hingga 59 bulan. Meskipun sebagian besar balita tergolong dalam kategori gizi normal. Namun, persentase balita yang terindikasi kekurangan gizi masih tergolong tinggi. Angka *stunting* tercatat sebesar 26,8%, *underweight* sebesar 23,9%, dan *wasting* sebesar 10,6%. Data ini menunjukkan bahwa permasalahan gizi, baik akut maupun kronis, masih menjadi isu kesehatan yang memerlukan perhatian khusus. Jika tidak segera ditangani, kondisi tersebut dapat berdampak negatif terhadap pertumbuhan fisik, perkembangan kecerdasan, serta produktivitas anak di masa depan.

Salah satu wilayah yang menghadapi tantangan serupa adalah Desa Sungai Ambangah. Terlihat dari ditemukannya status gizi tidak normal melalui pemantauan rutin Posyandu di Poskesdes. Temuan ini menegaskan perlunya metode klasifikasi gizi balita yang lebih akurat dan efisien sebagai langkah awal penanganan masalah gizi di perdesaan. Namun, Poskesdes Sungai Ambangah menghadapi tantangan dalam pengolahan data kesehatan, seperti nilai yang hilang, ketidakseimbangan antar kelas, serta keragaman data numerik dan kategorik, yang dapat memengaruhi akurasi model klasifikasi [2].

Dalam hal tersebut, di tengah perkembangan teknologi informasi dan ketersediaan data kesehatan yang semakin luas, pendekatan berbasis *data mining* menjadi potensi potensial [3]. Teknik *data mining* mampu mengidentifikasi pola tersembunyi dalam data, yang dapat digunakan untuk melakukan klasifikasi status gizi berdasarkan atribut seperti usia, berat badan, tinggi badan, jenis kelamin dan cara ukur. Berbagai Algoritma dapat digunakan untuk keperluan klasifikasi ini. Namun, dua metode yang cukup populer dan banyak digunakan yaitu Algoritma C4.5 dan *Naive Bayes* [4]. Algoritma C4.5 memiliki kapabilitas dalam membentuk pohon keputusan yang interpretatif serta mendukung pengolahan data bertipe numerik maupun kategorik [5]. Sementara itu, Algoritma *Naive Bayes* merupakan metode klasifikasi yang efisien dan bersifat sederhana karena mengandalkan prinsip dasar probabilitas dan statistik [6]. Meskipun keduanya banyak digunakan dalam berbagai penelitian, efektivitas masing-masing Algoritma dapat berbeda tergantung pada jenis dan karakteristik data yang digunakan.

Penelitian terkait penerapan algoritma C4.5 pada klasifikasi status gizi balita pernah di teliti oleh [7], tujuannya untuk mengidentifikasi faktor-faktor yang memengaruhi status gizi balita serta memperoleh hasil klasifikasi status gizi yang lebih akurat dan efisien. Melalui pengujian menggunakan aplikasi *RapidMiner*, diperoleh tingkat akurasi sebesar 95,50% dalam membedakan status gizi kategori baik dan buruk.

Selanjutnya, penelitian mengenai klasifikasi dengan menggunakan *data mining* untuk mendeteksi *stunting* pada balita menggunakan metode *Naive Bayes* dilakukan [8], penelitian ini menggunakan alat bantu *RapidMiner* serta atribut yang mempengaruhi pada penentuan *stunting* pada balita, yaitu umur, berat badan per umur (BB/U), dan tinggi badan per umur (TB/U). Hasil pengujian menunjukkan bahwa algoritma *Naïve Bayes* efektif dalam klasifikasi status *stunting* pada balita, dengan tingkat akurasi hingga mencapai 95,08%.

Berdasarkan permasalahan dan penelitian terkait, tujuan penelitian ini sebagai bahan uji banding tingkat akurasi metode Algoritma C4.5 dan *Naive Bayes* dalam mengklasifikasi status gizi pada balita. *Tools* atau *Software* yang digunakan untuk pengujian metode yaitu *RapidMiner*. Dataset yang digunakan yakni data yang didapatkan langsung dari Poskesdes Sungai Ambangah secara lengkap. Melalui penelitian ini, diharapkan memberikan informasi yang berharga bagi tenaga kesehatan dan peneliti dalam memilih algoritma yang sesuai untuk klasifikasi status gizi balita. Pemilihan algoritma yang tepat dapat membantu peningkatan akurasi klasifikasi, memungkinkan deteksi dini masalah gizi, perencanaan intervensi gizi yang lebih tepat sasaran, serta peningkatan kualitas layanan kesehatan dan status gizi balita secara keseluruhan.

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

II. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan teknik untuk mengidentifikasi pola dan ciri-ciri yang akan muncul di masa depan, serta untuk menemukan pengetahuan tak terduga yang sebelumnya belum pernah diketahui dari kumpulan data yang besar. Melalui proses *data mining*, dapat diperoleh pengetahuan baru yang sebelumnya belum terungkap [9] [10].

2.2. Algoritma C4.5

Algoritma C4.5, yang diperkenalkan oleh J. Ross Quinlan pada tahun 1993, merupakan evolusi dari algoritma sebelumnya yaitu ID3 (Iterative Dichotomiser). "Pohon keputusan ini dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama". Pengujian algoritma C4.5 diawali dengan memahami bagaimana metode ini bekerja dalam membentuk pohon keputusan berdasarkan nilai informasi dari atribut-atribut data. Proses dimulai dengan menghitung nilai *entropy*, yaitu ukuran tingkat ketidakpastian dari suatu himpunan data [11]. Nilai *entropy* dihitung menggunakan perhitungan berikut:

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

Keterangan:
S: himpunan kasus
n: jumlah partisi atribut A
Pi: proporsi data dalam kelas ke-i.

Selanjutnya, dilakukan perhitungan *information gain* untuk setiap atribut, yaitu pengukuran seberapa besar pengurangan *entropy* yang terjadi jika data dibagi berdasarkan atribut tersebut. *Gain* dihitung dengan rumus:

$$Rumus\ Gain(A) = Entropy\ (S) - \sum_{i=1}^{k} \frac{|Si|}{|S|} \times Entropy(Si)$$

Keterangan:
S: himpunan dataset (kasus)
A: atribut
k: jumlah partisi atribut A
|Si|: jumlah kasus pada partisi ke-i
|S|: jumlah kasus dalam S

2.3. Naïve Bayes

Naive Bayes adalah metode klasifikasi berbasis probabilistik yang sederhana, di mana prosesnya melibatkan perhitungan frekuensi serta penggabungan nilai-nilai dari data yang tersedia untuk menentukan kisaran probabilitas suatu kelas. Dalam pendekatannya, algoritma ini mengasumsikan bahwa setiap atribut dalam suatu kelas bersifat independen terhadap atribut lainnya. Ketika diterapkan pada dataset berukuran besar, *Naive Bayes* dikenal memiliki kecepatan pemrosesan dan tingkat akurasi yang cukup baik [11]. Pada algoritma *Naive Bayes*, pengujian dilakukan dengan pendekatan probabilistik berdasarkan *Teorema Bayes*. Berikut rumus yang digunakan:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Keterangan:
P(C|X) : Probalitas kelas C diberikan data X (disebut posterior)
P(X|C) : Probalitas data X muncuk jika kelasnya C (disebut likelihood)
P(C) : Probalitas awal kelas C sebelum melihat data (prior)
P(X) : Probalitas terjadinya data X

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

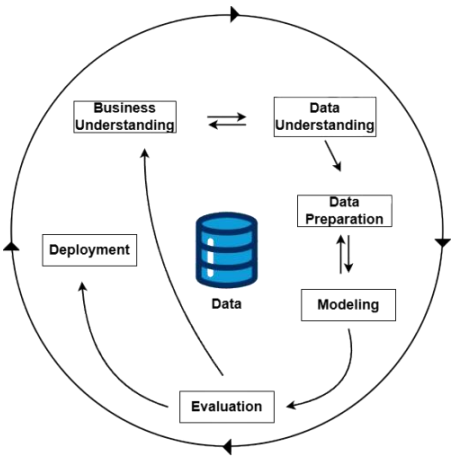
2.4. RapidMiner

RapidMiner adalah *software open-source* yang dikembangkan oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown and Raif Klinkenberg* dari rapid-i.com. Perangkat ini memiliki antarmuka grafis (GUI) yang memudahkan pengguna dalam pengoperasiannya. Dibangun menggunakan bahasa Java dan berlisensi GNU *Public License*, *RapidMiner* dapat dijalankan pada berbagai sistem operasi. Pengguna tidak memerlukan kemampuan pemrograman khusus karena seluruh fitur telah tersedia. RapidMiner dirancang khusus untuk keperluan pengolahan dan analisis *data mining* [12].

III. METODE PENELITIAN

3.1. Metode Penelitian

Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) diterapkan dalam penelitian ini sebagai kerangka sistematis untuk mengatur proses *data mining* secara terstruktur dan efisien[13].



Gambar 1. Tahapan Proses CRISP-DM

Dari gambar 1, menunjukkan bahwa CRIPS-DM terdiri dari enam tahapan proses dalam siklus hidupnya, yaitu:

1. *Business Understanding*

Tahap ini bertujuan merumuskan permasalahan serta tujuan dari penelitian terkait klasifikasi status gizi balita di Poskesdes Sungai Ambangah. Peneliti melakukan analisis masalah dan menetapkan ruang lingkup, serta mempelajari teori klasifikasi, khususnya algoritma C4.5 dan *Naive Bayes*, serta konsep *data mining* yang relevan.

2. *Data Understanding*

Pada tahap ini dilakukan eksplorasi terhadap data yang tersedia. Data primer diperoleh dari Poskesdes Desa Sungai Ambangah (Januari–Maret 2025) sebanyak 1.547 *record*. Eksplorasi dilakukan untuk memahami atribut, distribusi data, serta mengecek kelengkapan dan konsistensi.

3. *Data Preparation*

Dari 1.547 *record*, dibersihkan dengan menghapus 30 *record* tidak lengkap dan 15 *record noise*, menyisakan 1.502 data valid. Dari 17 atribut yaitu, Nama, Jenis Kelamin, Tanggal Lahir, Posyandu, Alamat, Umur Saat Ukur, Tanggal Ukur, Berat Badan saat ukur, Tinggi Badan saat ukur, Cara Ukur, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB, Naik Berat Badan. Kemudian, dipilih enam yang paling relevan seperti, Berat Badan, Tinggi Badan, Umur, Jenis Kelamin, Cara Ukur, dan BB/TB sebagai target klasifikasi. Data kemudian dikategorikan agar lebih mudah diproses oleh model.

4. *Modeling*

Pada tahap ini, penelitian menguji dua algoritma klasifikasi yaitu C4.5 dan *Naive Bayes*. Untuk algoritma C4.5, model dibangun dengan cara menjumlahkan nilai *entropy* dan *information gain* dari setiap atribut. Atribut dengan angka gain tertinggi dipilih sebagai *node* akar keputusan, kemudian proses dilanjutkan secara rekursif hingga membentuk sebuah struktur pohon keputusan lengkap yang berisi aturan-aturan klasifikasi berbentuk *if-then*. Sementara itu, pada algoritma *Naive Bayes*, model dibangun berdasarkan prinsip probabilitas. Proses diawali dengan menghitung probabilitas awal (*prior probability*)

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

dari masing-masing kelas, lalu menghitung probabilitas kemunculan setiap atribut pada kelas tertentu (*likelihood*). Probabilitas tersebut kemudian digabungkan menggunakan *teorema Bayes* untuk menghasilkan probabilitas posterior, dan kelas dengan nilai posterior tertinggi dipilih sebagai hasil prediksi.

5. *Evaluation*

Dalam penelitian ini, tahap evaluasi dilakukan dengan menggunakan teknik 10-*fold cross validation* untuk memastikan hasil yang stabil. Data dibagi menjadi sepuluh bagian, lalu secara bergantian sebagai bahan uji data dan data latih. Dari proses ini, performa algoritma diukur dengan metrik akurasi, presisi, *recall*, serta nilai AUC dari kurva ROC dengan lima kategori yaitu, *Fair Classification* (0,70–0,80), *Poor Classification* (0,60-0,70), dan *Failure* (0,50–0,60) [14]. Kemudian T-*Test* digunakan untuk menguji apakah terdapat perbedaan yang signifikan antara dua sampel yang tidak saling berhubungan berdasarkan nilai rata-rata hasil pengujiannya.

6. *Deployment*

Model klasifikasi status gizi yang dihasilkan dapat dimanfaatkan oleh tenaga kesehatan sebagai alat bantu dalam mengidentifikasi status gizi balita secara lebih efisien dan sistematis.

IV. HASIL DAN PEMBAHASAN

4.1. Data Understanding

Proses akumulasi data dilakukan oleh petugas kesehatan Poskesdes dalam bentuk *excel* yang di *input* dari pencatatan manual. Data yang diperoleh memerluan tahap pengolahan lebih lanjut sebelum dapat diubah menjadi informasi. Saat pengambilan data, tidak semua atribut digunakan untuk menjaga kerahasiaan data pribadi balita. Data yang dikumpulkan mencakup informasi gizi balita dari bulan Januari 2025 hingga Maret 2025, dengan total sebanyak 1.547 *record*. Berikut merupakan contoh data mentah yang diperoleh pada tabel 1.

Tabel 1. Data Balita Posyandu Desa Sungai Ambangah

JK	Usia	Berat (kg)	Tinggi (cm)	Cara Ukur	BB/U	TB/U
L	42	12.3	92.2	Berdiri	Normal	Pendek
L	13	9.2	78.9	Berdiri	Kurus	Pendek
L	43	12.4	88.6	Berdiri	Normal	Pendek
P	30	11.3	84	Berdiri	Normal	Pendek
P	18	9	75.9	Terlentang	Normal	Pendek
P	40	12.9	90.3	Berdiri	Normal	Pendek
L	57	12.6	90.4	Berdiri	Normal	Pendek
P	21	14.3	96	Berdiri	Normal	Pendek
L	31	12.3	92.2	Berdiri	Normal	Pendek
…	…	…	…	…	…	…
P	6	4.3	51.2	Terlentang	Kurus	Pendek

4.2. Data Preparation

Dalam pengolahan data sampel, dilakukan tahap data *cleaning*, *reduction transformation* untuk menyederhanakan dataset agar mudah dianalisis. Enam atribut utama yang digunakan adalah jenis kelamin, usia, cara ukur, BB/U, dan TB/U, dengan BB/TB sebagai label (*class*) prediksi status gizi balita, yang dikategorikan menjadi gizi normal dan gizi buruk. Penjelasan tiap atribut penting disajikan pada Tabel 2.

Tabel 2. Keterangan Atribut

No.	Atribut	Keterangan
1.	JK	Jenis Kelamin atau Gender (L/P)
2.	Usia (Bulan)	Usia balita yang dinyatakan dalam bulan
3.	Cara Ukur	Cara ukur yang dilakukan (terlentang/berdiri)
4.	BB/U	Berat badan balita berdasarkan umur yang dinyatakan dalam kilogram (kg)
5.	TB/U	Tinggi atau panjang badan balita berdasarkan umur yang dinyatakan dalam sentimeter (cm)
6.	BB/TB	Status gizi balita berdasarkan BB/TB

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

Pada atribut usia, berat badan dan tinggi badan dapat diklasifikasikan berdasarkan standar yang ditetapkan oleh WHO, yang digunakan sebagai acuan dalam penilaian status gizi balita [15]. Berikut rincian pembagian rentang usia balita berdasarkan kategori pada tabel 3.

Tabel 3. Kategori Atribut Usia

Usia	Kategori
< 12 bulan	Bayi
> 12 bulan	Anak

Kemudian, pembagian kategori berat badan dijabarkan pada tabel 4 berikut.

Tabel 4. Klasifikasi Atribut Berat Badan

Berat Badan	Kategori
< 2,5	Kurus
2,5 – 4	Normal
> 4	Gemuk

Klasifikasi atribut tinggi badan dalam penelitian dapat di kategorikan pada tabel 5 berikut.

Tabel 5. Klasifikasi Atribut Tinggi Badan

Tinggi Badan	Kategori
< 85	Pendek
85 – 110	Normal
>110	Tinggi

Berdasarkan tabel-tabel klasifikasi atribut yang telah dijelaskan, maka dataset awal pada tabel 2 dilakukan proses perubahan data dalam kategori tertentu yang lebih rumit dengan tidak menghitung isi, sehingga lebih sederhana untuk diolah. Berikut data *transformation* pada tabel 6 berikut.

Tabel 6. Tabel Transformation

No.	JK	Usia (Bulan)	Cara Ukur	BB/U	TB/U	BB/TB
1.	L	Anak	Berdiri	Normal	Pendek	Normal
2.	L	Anak	Berdiri	Kurus	Pendek	Gizi Buruk
3.	L	Anak	Berdiri	Normal	Pendek	Normal
4.	P	Anak	Berdiri	Normal	Pendek	Normal
5.	P	Anak	Terlentang	Normal	Pendek	Gizi Buruk
6.	P	Anak	Berdiri	Normal	Pendek	Normal
7.	L	Anak	Berdiri	Normal	Pendek	Normal
8.	P	Anak	Berdiri	Normal	Pendek	Normal
9.	L	Anak	Berdiri	Kurus	Pendek	Normal
…	.....	.....	.....	.....	.....	.....
1502.	P	Bayi	Terlentang	Kurus	Pendek	Gizi Buruk

4.3. Modeling

Pengujian dilakukan berdasarkan prinsip kerja algoritma C4.5 dan *Naive Bayes*. Kedua metode memiliki pendekatan berbeda saat mengolah data dan menghasilkan prediksi, sehingga pengujian disesuaikan dengan karakteristik masing-masing. Tujuannya adalah membandingkan performa kedua algoritma secara objektif dalam mengklasifikasi status gizi pada balita.

a. Uji Model Algoritma C4.5

Pengujian pertama menggunakan algoritma C4.5. Langkah awal meliputi pengolahan nilai *entropy* dan *gain* untuk menetapkan *node* awal (akar) pohon keputusan, dengan memilih atribut bernilai gain tertinggi sebagai *node* utama. Selanjutnya, perhitungan dilanjutkan untuk *node* turunan pertama (*node* 1.1), yang merupakan cabang dari *node* akar berdasarkan nilai atribut tertentu. Hasilnya ditampilkan pada tabel 7 (*node* 1) dan tabel 8 (*node* 1.1).

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

Tabel 7. Perhitungan Entropy dan Gain terhadap Node 1

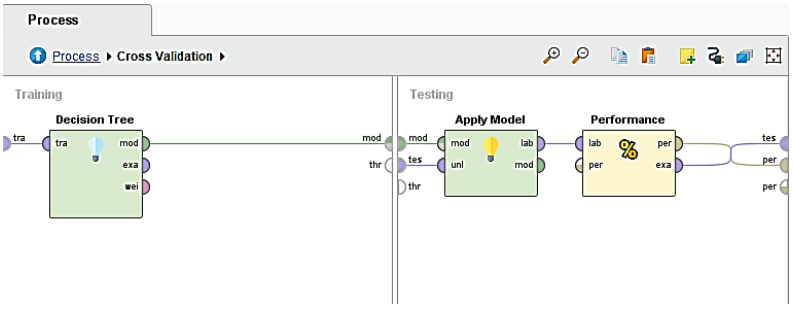
Node	Atribut	Kasus	Normal	Gizi Buruk	*Entropy*	*Gain*
1	Total	1502	1265	237	0.628995288	
	Jenis Kelamin					0.003242413
	Laki – laki	797	653	144	0.681557011	
	Perempuan	705	612	93	0.562666498	
	Usia					0.004133138
	Bayi	91	66	25	0.848163688	
	Anak	1411	1199	212	0.610460705	
	Cara Ukur					0.012328967
	Berdiri	1167	1014	153	0.56046197	
	Terlentang	335	251	84	0.812458791	
	Berat Badan					**0.40571834**
	Normal	1267	1234	33	0.174156351	
	Kurus	164	31	133	0.699426093	
	Gemuk	71	0	71	0	
	Tinggi Badan					0.107845774
	Normal	1357	1217	140	0.478960573	
	Pendek	145	48	97	0.915979809	

Nilai *entropy* dan gain yang tercantum pada Tabel 7 menunjukkan bahwa atribut Berat Badan memiliki nilai Gain terbesar, yaitu 0,40571834. Dengan demikian, atribut tersebut dipilih sebagai node pertama pada pohon keputusan (node 1). Proses dilanjutkan dengan perhitungan ulang *entropy* dan gain untuk menentukan pembagian selanjutnya (node 1.1), sebagaimana ditampilkan pada Tabel 8.

Tabel 8. Perhitungan Entropy dan Gain terhadap Node 1.1

Node	Atribut	Kasus	Normal	Gizi Buruk	*Entropy*	*Gain*
1.1	Total	1267	1234	33	0.628995288	
	Jenis Kelamin					0.004483181
	Laki – laki	663	638	25	0.231680064	
	Perempuan	604	596	8	0.101609312	
	Usia					0.001352781
	Bayi	67	63	4	0.326258815	
	Anak	1200	1171	29	0.164235652	
	Cara Ukur					0.012392336
	Berdiri	1002	988	14	0.106102129	
	Terlentang	265	246	19	0.372228961	
	Tinggi Badan					**0.066533842**
	Normal	1195	1187	8	0.057979247	
	Pendek	72	47	25	0.931562768	

Perhitungan entropy dan gain pada Tabel 8 menunjukkan bahwa atribut Tinggi Badan memiliki nilai Gain tertinggi sebesar 0,066533842 dalam penentuan node 1.1. Oleh karena itu, atribut tersebut dipilih sebagai node 1.1 pada model pohon keputusan. Setelah pengujian awal dilakukan melalui Microsoft Excel, validasi lanjutan dilakukan menggunakan perangkat lunak RapidMiner dengan metode *10-fold cross validation*, sebagaimana diilustrasikan pada Gambar 2



Gambar 2. Sub Proses Training dan Testing Algoritma C4.5

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

Berdasarkan gambar 2 menunjukkan pada sisi *training*, ditujukan untuk membentuk model menggunakan algoritma C4.5 (*Decision Tree*). Selanjutnya, pada bagian *Testing*, model yang terbentuk kemudian diterapkan di data uji menggunakan operator *Apply Model*, lalu dievaluasi menggunakan operator *Performance*. Hasil evaluasi menunjukkan tingkat keakuratan model dalam mengklasifikasikan status gizi, yang ditampilkan pada gambar 3.

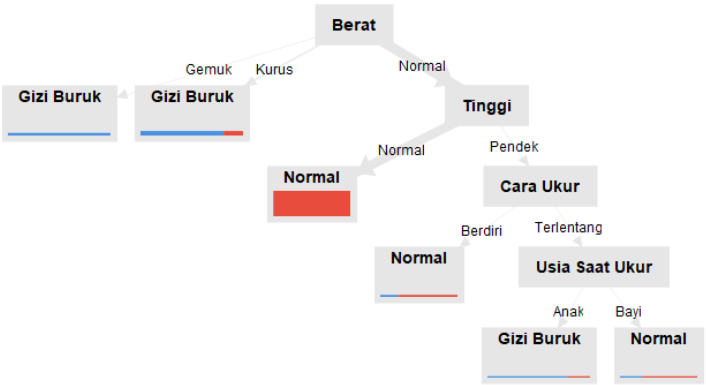
accuracy: 96.01% +/- 1.24% (micro average: 96.01%)

	true Gizi Buruk	true Normal	class precision
pred. Gizi Buruk	210	33	86.42%
pred. Normal	27	1232	97.86%
class recall	88.61%	97.39%	

Gambar 3. Confusion Matrix Algoritma C4.5

Berdasarkan gambar 3 menunjukkan hasil pengujian 451 data uji, di mana dari 1.502 data, sebanyak 210 data diprediksi sesuai sebagai kelas "Gizi buruk" dan 33 data diprediksi gizi buruk ternyata termasuk dalam kelas "Normal". Kemudian, sebanyak 27 data di prediksi "Normal" ternyata termasuk dalam kelas "Gizi buruk". Di sisi lain, terdapat 1.232 data "Normal" diprediksi dengan benar, serta terdapat 27 data yang diprediksi "Normal" tetapi ternyata termasuk dalam kategori "Gizi Buruk". Model menghasilkan akurasi keseluruhan sebesar 96.01%, yang merupakan rata-rata dari 10-*fold* serta nilai AUC sebesar 0.972 teridentifikasi bahwa model memunyai kemampuan sangat baik pada saat membedakan antara kategori kelas normal serta gizi buruk.

C4.5 juga menghasilkan pohon keputusan berdasarkan atribut seperti berat badan, tinggi badan, cara ukur, dan usia. Setiap simpul menunjukkan kondisi atribut, dengan hasil akhir klasifikasi Gizi Buruk atau Normal. Visualisasi pohon ditampilkan pada gambar berikut.



Gambar 4. Pohon Keputusan (Decision Tree)

Gambar tersebut menampilkan pohon keputusan hasil algoritma C4.5 untuk memilah status dari gizi balita. Pohon dimulai dari atribut Berat, lalu bercabang ke Tinggi, Cara Ukur, dan Usia Saat Ukur. Setiap cabang menunjukkan alur pengambilan keputusan hingga mencapai simpul daun yang menyatakan klasifikasi Gizi Buruk atau Normal. Struktur ini memperlihatkan pola keputusan berdasarkan kombinasi atribut yang membedakan status gizi balita.

b. Uji Model Naïve Bayes

Perhitungan selanjutnya dilakukan menggunakan algoritma *Naive Bayes*. Langkah awal proses adalah melakukan perhitungan nilai probabilitas *prior* untuk setiap kelas, yaitu gizi normal dan gizi buruk. Hasil perhitungannya dapat di lihat pada tabel 9.

Tabel 9. Perhitungan Algoritma Naïve Bayes

Atribut	Kasus	Normal	Gizi Buruk	P(X|Ci)	
				Normal	Gizi Buruk
Total	1502	1265	237	0.842210386	0.157789614
Jenis Kelamin					

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

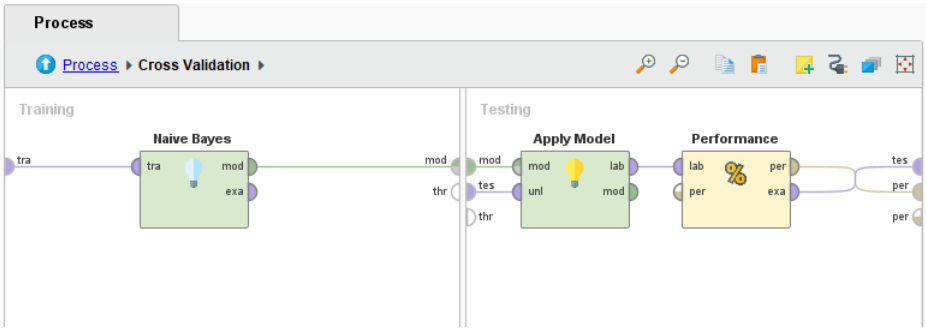
Atribut	Kasus	Normal	Gizi Buruk	P(X|Ci)	
				Normal	Gizi Buruk
Laki – laki	797	653	144	0.516205534	0.607594937
Perempuan	705	612	93	0.483794466	0.392405063
Usia					
Bayi	91	66	25	0.052173913	0.105485232
Anak	1411	1199	212	0.947826087	0.894514768
Cara Ukur					
Berdiri	1167	1014	153	0.801581028	0.64556962
Terlentang	335	251	84	0.198418972	0.35443038
Berat Badan					
Normal	1267	1234	33	0.975494071	0.139240506
Kurus	164	31	133	0.024505929	0.561181435
Gemuk	71	0	71	0	0.299578059
Tinggi Badan					
Normal	1357	1217	140	0.962055336	0.5907173
Pendek	145	48	97	0.037944664	0.4092827

Pada tabel 9 diperoleh hasil perhitungan probabilitas dari setiap atribut terhadap kelas "Gizi Normal" dan "Gizi Buruk". Kemudian, pada tabel 10 terdapat data X yang digunakan sebagai studi kasus baru. Data ini dihitung hingga menggunakan rumus *Naive Bayes* hingga diperoleh nilai probabilitas tertinggi, yang akan menjadi hasil klasifikasi gizi balita. Berikut hasil perhitungannya:

Tabel 10. Perhitungan nilai probalitas prior untuk data X

Atribut	P(X|Ci)	
	Normal	Gizi Buruk
Prior	0.842210386	0.10918775
Jenis Kelamin		
Laki-laki	0.516205534	0.609756098
Usia		
Anak	0.947826087	0.884146341
Cara Ukur		
Berdiri	0.801581028	0.609756098
Berat Badan		
Kurus	0.024505929	0.634146341
Tinggi Badan		
Pendek	0.037944664	0.591463415
Nilai Probabilitas	0.000307143	**0.013462563**

Berdasarkan hasil perhitungan pada tabel 10, diperoleh nilai probabilitas akhir untuk kategori masing-masing kelas. Nilai probabilitas tertinggi terdapat pada kelas "Gizi Kurang" dengan nilai sebesar 0.013462563. Dari studi kasus tersebut dapat disimpukan bahwa balita diprediksi memiliki status Gizi Buruk. Tahapan verifikasi hasil perhitungan yang sebelumnya dilakukan menggunakan Microsoft Excel dilanjutkan dengan pengujian menggunakan perangkat lunak RapidMiner melalui operator *cross validation*. Ilustrasi proses pengujian tersebut ditunjukkan pada Gambar 5



Gambar 5. Sub Proses Training dan Testing Algoritma Naive Bayes

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

Berdasarkan pada gambar 5, memperlihatkan alur proses validasi model *Naive Bayes*. Pada uji tahap *training*, data latih diaplikasikan untuk membentuk suatu model menggunakan operator *Naive Bayes*. Setelah model terbentuk, tahap testing dilakukan dengan menerapkan model tersebut pada data uji menggunakan operator *Apply Model*. Hasil prediksi kemudian dievaluasi menggunakan operator *Performance* untuk mengetahui seberapa baik model dalam mengklasifikasikan data, berdasarkan metrik seperti akurasi, presisi, dan *recall* yang terlihat pada gambar 6 berikut.

accuracy: 95.54% +/- 1.29% (micro average: 95.54%)

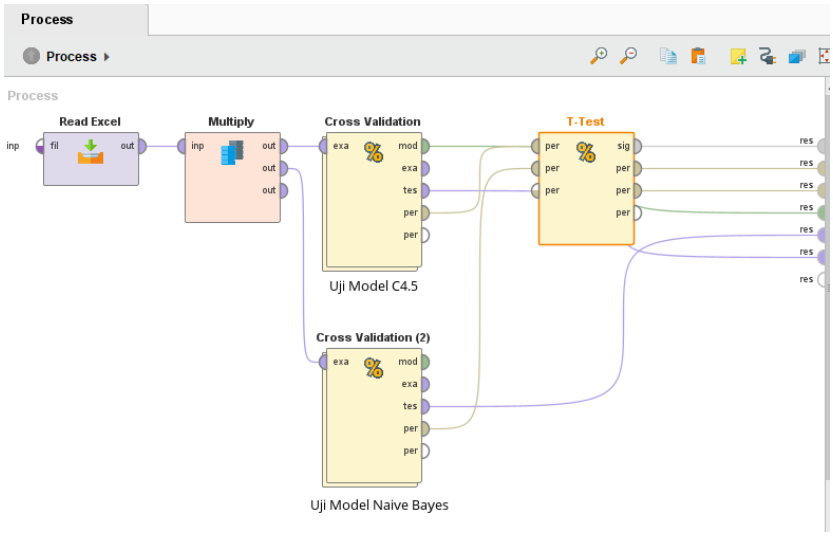
	true Gizi Buruk	true Normal	class precision
pred. Gizi Buruk	207	37	84.84%
pred. Normal	30	1228	97.62%
class recall	87.34%	97.08%	

Gambar 6. Confusion Matrix Algoritma Naive Bayes

Berdasarkan gambar 6, dari hasil evaluasi diketahui bahwa 1.502 data terdapat 207 data yang benar diprediksi sebagai "Gizi Buruk", dan 1.228 data yang benar diprediksi sebagai "Normal". Namun, terdapat 30 data yang sebenarnya termasuk dalam kelas "Gizi Buruk" tetapi diprediksi sebagai Normal, dan 37 data yang sebenarnya "Normal" tetapi diprediksi sebagai "Gizi Buruk". Model ini menghasilkan akurasi keseluruhan sebesar 95.54% serta nilai AUC sebesar 0.983, yang menunjukkan performa klasifikasi sangat baik.

4.4. Evaluation

Analisa perbandingan algoritma dilakukan dengan menggunakan perangkat lunak *RapidMiner* untuk menguji sejauh mana perbedaan performa antara dua algoritma yang di analisis. Salah satu metode statistik yang diterapkan pada proses ini adalah Uji T-*Test*, dengan tujuan untuk mengetahui apakah memiliki perbedaan yang signifikan antara dua sampel yang tidak berhubungan berdasarkan nilai rata-rata hasil pengujiannya. Berikut proses Uji T-*Test* menggunakan *RapidMiner* yang menunjukkan model visual dari pengujian tersebut dalam pemrosesan data pada gambar 7.



Gambar 7. Confusion Matrix Algoritma Naive Bayes

Pada gambar 7, menunjukkan alur proses analisis data yang diawali dengan *Read Excel* untuk membaca dataset. Data kemudian diduplikasi menggunakan *Multiply*, sehingga menghasilkan dua salinan dataset untuk evaluasi model yang berbeda. Salinan pertama digunakan untuk proses *Cross Validation* dengan Algoritma C4.5 (*Decision Tree*), sedangkan salinan kedua digunakan untuk *Cross Validation*(2) dengan algoritma *Naive Bayes*. Masing-masing proses validasi melakukan pelatihan (*training*) dan pengujian (testing) model sebanyak 10 kali lipat (10-fold). Hasil evaluasi dari kedua model kemudian

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

dimasukkan ke dalam operator T-*Test* untuk membandingkan performa kedua algoritma secara statistik seperti gambar 8 berikut.

A	B	C
	C4.5	Naive Bayes
C4.5		0.479
Naive Bayes		

Gambar 8. Confusion Matrix Algoritma Naive Bayes

Penelitian ini menggunakan tingkat signifikansi (α) sebesar 0,05. Jika nilai *p-value* melebihi 0,05, maka H0 diterima, yang mengindikasikan tidak adanya perbedaan signifikan. Namun, jika *p-value* berada di bawah 0,05, maka H0 ditolak karena terdapat perbedaan signifikan. Berdasarkan hasil pada Gambar III.7, nilai *p-value* antara model C4.5 dan Naive Bayes adalah 0,479, yang lebih tinggi dari nilai α. Hal ini menunjukkan bahwa secara statistik, tidak terdapat perbedaan signifikan antara performa kedua algoritma. Ringkasan hasil perbandingan disajikan pada tabel berikut.

Tabel 11. Perhitungan nilai probalitas prior untuk data X

	C4.5	***Naive Bayes***
Accuracy	96.01%	95.54%
***Area Under Curve* (AUC)**	0.972	0.983
Precision	Gizi Buruk : 86,42% Normal : 97,86%	Gizi Buruk : 84,84% Normal : 97,62%
Recall	Gizi Buruk : 88,61% Normal : 97,39%	Gizi Buruk : 87,34% Normal : 97,08%

Berdasarkan tabel 11, perbandingan berdasarkan tingkat *accuracy* dalam menentukan klasifikasi status gizi pada Poskesdes Desa Sungai Ambangah, Kalimantan Barat dengan 1502 dataset yang terdiri dari enam atribut yaitu Jenis Kelamin, Usia, Cara Ukur, Berat Badan, Tinggi Badan, dan BB/TB sebagai label (*class*). Hasil evaluasi menunjukkan bahwa algoritma C4.5 memiliki performa yang lebih baik dengan hasil *accuracy* 96.01% dan nilai AUC sebesar 0.972. Sementara itu, algoritma *Naive Bayes* menghasilkan *accuracy* 95.54% dan AUC sebesar 0.983. Dari hasil pengujian, berdasarkan kategori *performance* AUC secara keseluruhan kedua algoritma masuk dalam kategori *Excellent Classification* (rentang 0,90–1,00), yang menunjukkan bahwa keduanya mampu mengklasifikasikan status gizi balita dengan sangat baik.

V. KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa *data mining* efektif untuk klasifikasi status gizi balita dengan algoritma C4.5 dan *Naive Bayes*. Tingkat efektivitas kedua algoritma diukur menggunakan metrik evaluasi kinerja model, yaitu akurasi dan Area Under Curve (AUC). Secara numerik, C4.5 memiliki akurasi lebih tinggi yaitu 96,01%, sementara *Naive Bayes* unggul pada AUC 0,983. Untuk mengetahui apakah perbedaan kinerja kedua algoritma tersebut signifikan secara statistis, dilakukan uji T-*Test*, yang menghasilkan nilai p-*value* 0,479 (> 0,05), sehingga dapat disimpulkan bahwa tidak terdapat perbedaan yang signifikan antara kedua algoritma. Meskipun demikian, C4.5 lebih direkomendasikan untuk diimplementasikan karena menghasilkan pohon keputusan yang mudah dipahami sehingga dapat mendukung proses pengambilan keputusan secara praktis di lapangan.

Berdasarkan hasil penelitian ini, disarankan untuk menambah atribut relevan seperti Riwayat penyakit, pola makan, imunisasi, dan sosial ekonomi, serta memperbesar jumlah data dengan rentang waktu lebih Panjang untuk meningkatkan akurasi dan generalisasi model. Penelitian selanjutnya dapat membandingkan algoritma lain seperti SVM, *Random Forest*, dan *Neural Networks* untuk memperoleh algoritma klasifikasi status gizi balita yang lebih optimal. Selain itu, model klasifikasi ini berpotensi di kembangkan menjadi aplikasi *web* atau *mobile* sebagai sistem pendukung keputusan, sehingga dapat digunakan kader Posyandu atau bidan desa dalam mengidentifikasi status gizi balita secara praktis.

DOI: https://doi.org/10.35457/antivirus.v20i1.5245

DAFTAR RUJUKAN

[1] Ummah, M. S., "Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka". Dalam Sustainability (Switzerland), vol. 11, nomor 1, 2025.

[2] Y. Rifa'i, "Analisis Metodologi Penelitian Kulitatif dalam Pengumpulan Data di Penelitian Ilmiah pada Penyusunan Mini Riset," *Cendekia Inovatif Dan Berbudaya*, vol. 1, no. 1, pp. 31–37, 2023, doi: 10.59996/cendib.v1i1.155.

[3] K. L. Kohsasih and Z. Situmorang, "Comparative Analysis of C4.5 and Naïve Bayes Algorithms in Predicting Cerebrovascular Disease," *Jurnal Informatika*, vol. 9, no. 1, pp. 13–17, 2022.

[4] A. Lasarudin, Hilmansyah Gani, and Misran Tomayahu, "Perbandingan Metode Naïve Bayes dan C4.5 Klasifikasi Status Gizi Bayi Balita," *SPECTA Journal of Technology*, vol. 6, no. 3, pp. 273–283, 2023, doi: 10.35718/specta.v6i3.789.

[5] T. R. Matondang, Y. Ramadhan Nasution, Armansyah, and M. Furqan, "Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita," *Jurnal Fasilkom*, vol. 14, no. 1, pp. 216–225, 2024, doi: 10.37859/jf.v14i1.6941.

[6] K. Handayani, L. Lisnawanty, A. Latif, M. R. Firdaus, and F. N. Hasan, "Komparasi Algoritma C4.5 Dan Naïve Bayes Dalam Penentuan Status Kelayakan Donor Darah," *Sistemasi*, vol. 10, no. 3, p. 676, 2021, doi: 10.32520/stmsi.v10i3.1440.

[7] R. A. A. Sri Lestari, "Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat," *Jurnal Wahana Pendidikan*, vol. 8, no. 10, pp. 116–125, 2022.

[8] Y. Mulyanto, F. Idifitriani, A. Wati, U. T. Sumbawa, D. Mining, and K. P. Tano, "Vol 7 No 2 , September 2024 KLASIFIKASI DATA MINING UNTUK PENENTUAN STUNTING," vol. 7, no. 2, pp. 119–125, 2024.

[9] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 2, 2021, doi: 10.31294/p.v23i2.11395.

[10] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.

[11] J. N. Permana, R. Goejantoro, and S. Prangga, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Prediksi Ketepatan Waktu Studi Mahasiswa (Studi Kasus: Program Studi Statistika Universitas Mulawarman)," *Jurnal EKSPONENSIAL*, vol. 13, pp. 161–170, 2022, [Online]. Available: https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/947

[12] P. Meilina, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan C4.5," *Teknologi*, vol. 9, no. 4, pp. 2862–2869, 2022.

[13] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.

[14] A. Nur, A. Rohim, A. I. Purnamasari, and I. Ali, "Program Keluarga Harapan (Studi Kasus : Kecamatan Cicalengka Kabupaten Bandung)," vol. 8, no. 2, pp. 2355–2362, 2024.

[15] M. Y. Titimeidara and W. Hadikurniawati, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," *Jurnal Ilmiah Informatika*, vol. 9, no. 01, pp. 54–59, 2021, doi: 10.33884/jif.v9i01.3741.

DOI: https://doi.org/10.35457/antivirus.v20i1.5245