

PENERAPAN ALGORITMA C5.0 DALAM KLASIFIKASI FAKTOR-FAKTOR PENYEBAB TERJADINYA DBD

St Syahdan ⁽¹⁾, Rospina Rura ⁽²⁾, Ratna Dwi Christyanti⁽³⁾

^{1,2,3}Jurusan Matematika, Universitas Kaltara, Bulungan

E-mail: stsyahdan89@gmail.com

ABSTRACT

Dengue hemorrhagic fever (DHF) is a disease caused by the dengue virus and transmitted by the Aedes aegypti mosquito. In fact, in 2019 the Bulungan District Health Office determined the findings of a DHF case in Bulungan as an extraordinary incident. The purpose of this study was to determine the variables that influence the factors that cause DHF using the C5.0 Algorithm method. One method that can be used to analyze which factors most influence the factors that cause DHF is using the C5.0 Algorithm. The results of data analysis show that the most influencing variables in analyzing the factors of DHF occurrence in Bulungan Regency are age, history of disease, history of co-morbidities, gender, and history of DHF. Based on the results of the analysis and discussion of the classification of the data on the factors that cause DHF, the following conclusions can be drawn: based on the analysis of the C5.0 algorithm, there are five variables including age, gender, history of DHF, history of co-morbidities, and employment status. Of the five variables, the age variable is the attribute that makes the biggest contribution which can be concluded that from the five variables it can be obtained that the factors that significantly influence the occurrence of DHF in Bulungan Regency are age, sex, history of co-morbidities and history of DHF.

Keywords : *Algorithm C5.0, Classification, decision tree.*

ABSTRAK

Demam berdarah *dengue* (DBD) merupakan penyakit yang disebabkan oleh virus *dengue* dan ditular oleh nyamuk *Aedes Aegypti*. Pada tahun 2019 Dinas Kesehatan Kabupaten Bulungan menetapkan temuan kasus DBD di Bulungan sebagai kejadian luar biasa. Tujuan dari penelitian ini adalah untuk mengetahui variabel yang mempengaruhi faktor-faktor penyebab terjadinya penyakit DBD menggunakan metode *Algoritma C5.0*. Salah satu metode yang dapat digunakan untuk menganalisis faktor mana yang paling mempengaruhi penyebab terjadinya DBD menggunakan *Algoritma C5.0*. Berdasarkan hasil analisis dan pembahasan klasifikasi dari data faktor-faktor penyebab terjadinya DBD dapat disimpulkan sebagai berikut: berdasarkan analisis algoritma C5.0 terdapat lima variabel diantaranya variabel umur, jenis kelamin, riwayat pernah DBD, riwayat penyakit penyerta, serta status pekerjaan. Dari kelima variabel tersebut variabel umur merupakan atribut yang memberikan kontribusi terbesar sehingga disimpulkan bahwa dari kelima variabel tersebut dapat diperoleh bahwa faktor-faktor yang mempengaruhi terjadinya DBD secara signifikan di Kabupaten Bulungan adalah variabel umur, jenis kelamin, riwayat penyerta serta riwayat pernah DBD.

Kata kunci: *Algoritma C5.0, Klasifikasi, Pohon Keputusan.*

Article History

Received : 6 Juni 2024
Accepted : 18 Juni 2024

Revised : 10 Juni 2024
Published : 30 Juni 2024

Sitasi:

Syahdan, S., Rura, R., Christyanti, R. D. (2024) Penerapan Algoritma C5.0 dalam Klasifikasi Faktor-Faktor Penyebab Terjadinya DBD. Jurnal Sains Benuanta, 3(2), 1–7. DOI: <https://doi.org/10.61323/jsb.v3i1.105>

1. Pendahuluan

Demam Berdarah *Dengue* (DBD) merupakan penyakit yang disebabkan oleh virus *dengue* dan ditular oleh nyamuk *Aedes Aegypti*. DBD salah satu penyakit menular yang masih menjadi masalah kesehatan Indonesia, karena berisiko menyebabkan kematian serta penyebaran yang sangat cepat. Berdasarkan data yang di input Widyawati (2021), saat ini kasus DBD terbesar di 472 kabupaten/kota di 34 provinsi. Di Kabupaten Bulungan sendiri kasus DBD termasuk kasus terbesar di bidang kesehatan di kabupaten bulungan. Berdasarkan Velix kepada harian rakyat Kaltara (2022), tercapai 353 kasus DBD sepanjang tahun 2022 kasus tersebut meningkat dari tahun sebelumnya bila di bandingkan pada tahun 2021 tercatat 73 kasus dan 1 orang meninggal dunia, hingga di ketahui kasus tersebut meningkat tiga kali lipat pada tahun 2022.

Perkembangbiakan jentik nyamuk *Aedes* penyebab DBD dipengaruhi oleh beberapa faktor seperti iklim, curah hujan, suhu, kelembaban dan faktor geografis. Faktor lain yang mempengaruhi diantaranya umur, tingkat pendidikan, perilaku host terhadap kejadian demam berdarah dengue (DBD), kemiskinan, dan pekerjaan (Respati et al., 2017). Menurut (Novrita et al., 2017) berdasarkan faktor umur, kekebalan tubuh yang rendah pada anak-anak membuat mereka lebih rentan terkena DBD sedangkan pada faktor pekerjaan, status pekerja juga lebih rentan terkena DBD dikarenakan mereka banyak menghabiskan waktu di luar ruangan dimana nyamuk *Aedes Aegypti* mencari mangsa. Kebanyakan anak yang mengalami asma bronchial beresiko untuk mengalami kematian (Hikmah & Kasmini H, 2015). Selain itu kepadatan tempat tinggal dan lingkungan yang tidak bersih serta tergenang air merupakan tempat berkembangbiaknya larva (Marwanti & Wahyono, 2019).

Dalam mengatasi masalah DBD dapat di selesaikan dengan analisis Algoritma C5.0 Algoritma C5.0 merupakan suatu teknik yang dapat menganalisis data menjadi sekumpulan aturan yang nantinya dapat dijadikan masukan dalam pengambilan keputusan dengan berdasarkan Algoritma decision tree atau pohon keputusan. Pohon keputusan merupakan pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan (Pratiwi et al., 2020). Selain itu, pohon keputusan (*decision tree*) merupakan suatu model prediksi yang pembangunannya relative cepat serta mudah untuk di pahami (Ginting et al., 2014). Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*), simpul pada sebuah pohon di bedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan/internal (*branch/internal node*) dan simpul daun (*leaf node*) (Latifah et al., 2018; Sulaiman et al., 2021). Menurut Prasetyo (2014) terdapat tiga kelompok penting dalam syarat pengujian *node* yaitu (1) fitur biner, (2) fitur bertipe kategorial, dan (3) fitur bertipe numerik. Algoritma C5.0 adalah salah satu Algoritma *data mining* yang khususnya diterapkan pada Algoritma *decision tree*. Algoritma C5.0 ini merupakan penyempurnaan Algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Dalam Algoritma ini pemilihan atribut diproses menggunakan *gain ratio*. Dalam penelitian yang dilakukan oleh Amalda (2022) peneliti melakukan implementasi Algoritma C5.0 dalam menganalisis kelayakan penerima keringanan UKT Mahasiswa IKT, kesimpulan dari hasil peneliti tersebut terdapat 108 data Mahasiswa IKT yang mengajukan Banding UKT di peroleh akurasi sebesar 91% pada sistem pengambilan keputusan, sedangkan pada sistem penentuan UKT diperoleh akurasi sebesar 80%. Sedangkan

penelitian yang dilakukan oleh Rahmayanti dkk (2020) dengan judul implementasi Algoritma C5.0 untuk menganalisis gejala prioritas pada anak yang mengalami *bullying* dan hasil kesimpulan diperoleh Algoritma C5.0 mampu melakukan seleksi fitur sebesar 93,33% lebih baik dibandingkan dengan metode Algoritma *Naive Bayes* dan KNN dengan hasil akurasi sebesar 92,77%.

2. Metode

Penelitian ini merupakan penelitian terapan yang bertujuan untuk menerapkan, menguji dan mengevaluasi kemampuan teori yang diterapkan dalam memecahkan masalah-masalah praktis. Jenis data yang digunakan dalam penelitian ini adalah data sekunder yaitu data yang diperoleh dari data rekam medis di RSUD Dr. H. Soemarno Sosroatmodjo Kabupaten Bulungan yang memuat keadaan pasien umum. Populasi dalam penelitian ini merupakan jumlah pasien umum yang ada di RSUD. Dr.H. Soemarno Sosroatmodjo Kabupaten Bulungan. Adapun sampel dalam penelitian ini adalah pasien yang tercatat direkam medis.

Prosedur yang dilakukan dalam menyelesaikan permasalahan dan mencapai tujuan adalah (1) mengidentifikasi dan merumuskan masalah (2) mencari berbagai referensi sebagai sumber informasi melalui beberapa literatur terkait penerapan algoritma C5.0 (3) mengumpulkan dan menganalisis data, (4) terakhir membuat kesimpulan dari hasil klasifikasi faktor-faktor penyebab terjadinya DBD yang diperoleh.

Langkah kerja pembuatan *tree* pada Algoritma C5.0 mirip dengan pembuatan *tree* pada Algoritma C4.5 kemiripan tersebut meliputi perhitungan *entropy* dan *gain*. Jika pada Algoritma C4.5 berhenti sampai perhitungan *gain*, maka pada Algoritma C5.0 akan melanjutkannya dengan menghitung *gain ratio* dengan menggunakan *gain* dan *entropy* yang telah ada.

Rumus yang digunakan untuk mencari nilai *entropy* dapat dilihat pada persamaan 1 (Chair et al., 2017)

$$Entropy (s) = - \sum_{i=1}^n p_i^2 \log p_i \quad \text{Pers. 1}$$

Keterangan:

- S = Himpunan kasus
- n = Jumlah partisi
- p_i = Proporsi dari S_i terhadap S
Dimana $i = 1, 2, \dots, n$

Selanjutnya nilai *gain* dapat dihitung dengan menggunakan persamaan 2 (Latifah et al., 2018):

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

Keterangan:

- S = Himpunan kasus
- A = Atribut
- n = Jumlah partisi
- $|S_i|$ = Jumlah kasus pada partisi ke- i
- $|S|$ = Jumlah kasus dalam S

Setelah didapat nilai *entropy* dan *gain*, langkah selanjutnya adalah menghitung nilai *gain ratio*. Adapun rumus dari perhitungan *gain ratio* adalah sebagai berikut (Asroni, dkk 2018):

$$Gain\ ratio = \frac{Gain (S, A)}{\sum_{i=1}^n Entropy (S_i)} \quad \text{Pers. 3}$$

Keterangan:

$Gain(S, A)$ = Nilai *gain* dari satu variabel

$\sum_{i=1}^n Entropy(S_i)$ = Jumlah nilai *entropy* dalam suatu variabel

Proses diulang untuk masing-masing cabang sampai semua kelas pada cabang memiliki kelasnya masing-masing.

3. Hasil dan Pembahasan

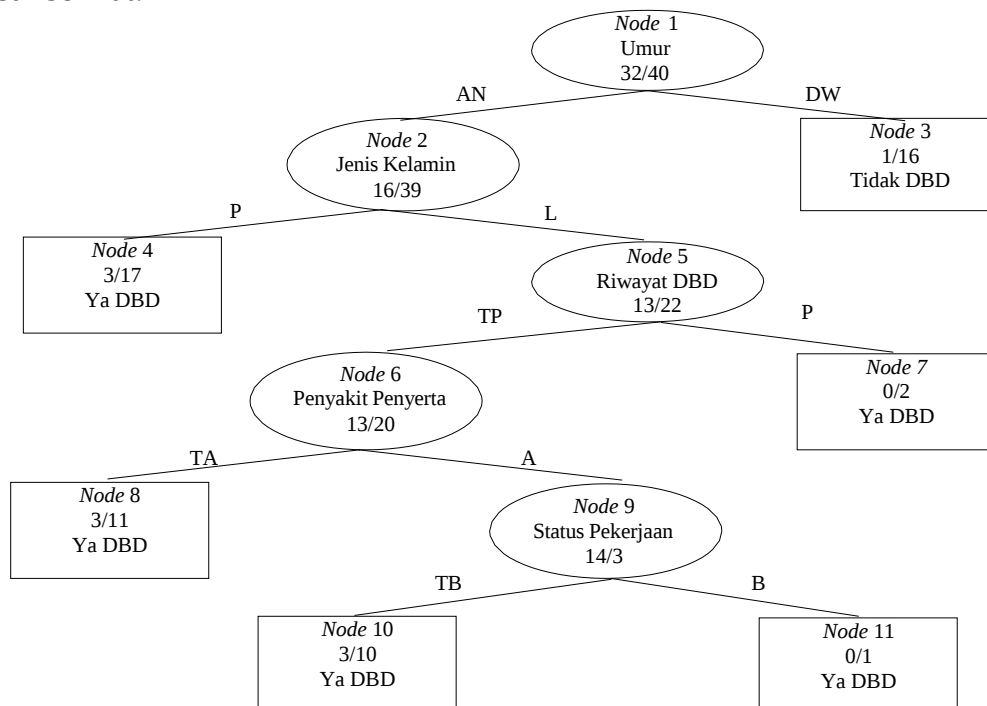
Data yang digunakan dalam penelitian ini adalah data pasien yang tercatat direkam medis DBD di RSUD Dr. H. Soemarno Sosroatmodjo Kabupaten Bulungan yang diperoleh dari jurnal ilmiah yang ditulis oleh Syahdan dkk (2021) dengan judul *Analysis of the Factors that Cause Dengue Hemorrhagic Fever (DHF) Using Chi-Square Automatic Interaction Detection (CHAID)* yaitu terdapat 72 pasien. Dari data tersebut diketahui bahwa untuk faktor jenis kelamin (JK) terdiri dari dua kategori yaitu perempuan (P) sebanyak 24 pasien dan laki-laki (L) sebanyak 48 pasien. Untuk kategori perempuan terdapat 7 yang tidak DBD dan 17 yang DBD sedangkan untuk kategori laki-laki terdapat 26 yang tidak DBD dan 22 yang DBD. Untuk faktor umur yang terdiri dari dua kategori yaitu anak (AN) sebanyak 55 pasien dan dewasa (DW) sebanyak 17 pasien. Untuk kategori anak terdapat 17 yang tidak DBD dan 38 yang DBD sedangkan untuk kategori dewasa terdapat 16 tidak DBD dan 1 yang DBD. Faktor riwayat pernah DBD yang terdiri dari dua kategori yaitu tidak pernah DBD (TP) sebanyak 70 pasien dan pernah DBD (P) sebanyak 2 pasien. Untuk kategori tidak pernah DBD terdapat 33 pasien yang tidak DBD dan 37 yang DBD sedangkan untuk kategori pernah DBD terdapat 0 yang tidak DBD dan 2 yang DBD. Faktor riwayat penyakit penyerta dengan 2 kategori yaitu tidak ada riwayat penyakit penyerta (TA) sebanyak 30 pasien dan ada riwayat penyakit penyerta sebanyak 42 pasien. Untuk kategori tidak ada riwayat penyakit terdapat 10 yang tidak DBD dan 20 yang DBD sedangkan untuk kategori ada riwayat penyakit terdapat 23 yang tidak DBD dan 19 yang DBD. Selanjutnya untuk faktor status pekerjaan dengan 2 kategori yaitu tidak bekerja (TB) sebanyak 61 pasien dan kategori bekerja sebanyak 11 pasien. Untuk kategori yang tidak bekerja terdapat 22 yang tidak DBD dan 39 yang DBD sedangkan untuk kategori bekerja terdapat 11 yang tidak DBD dan 0 yang DBD.

Pada proses pembentukan pohon klasifikasi Algoritma C5.0 dilakukan analisis untuk mencari nilai *entropy*, nilai *gain* dan nilai *gain ratio* pada masing-masing variabel. Pada tahap pertama yaitu menentukan *node* akar berdasarkan nilai *gain* tertinggi dari setiap variabel yang ada. Sebelum menghitung nilai *gain* setiap variabel, nilai *entropy* total untuk seluruh data harus cari terlebih dahulu dengan menggunakan pers. 1. Dari hasil perhitungan diperoleh nilai *entropy* total sebesar 0,991076. Nilai *entropy* total ini selanjutnya akan digunakan pada setiap variabel untuk menghitung nilai *entropy* dengan menggunakan pers. 1, nilai *gain* dengan menggunakan pers. 2 dan nilai *gain ratio* dengan menggunakan pers. 3. Hasil perhitungan nilai *gain ratio* untuk semua variabel bebas dapat dilihat pada Tabel berikut

Tabel 1 Hasil Perhitungan *Entropy Gain* dan *Gain Ratio* Untuk *Node* Akar.

<i>Node</i>	Variabel	Jumlah Kasus	TD	YD	<i>Entropy</i>	<i>Gain</i>	<i>Gain Ratio</i>
	Total	72	32	40	0.991076		
1	X_1 P	24	7	17	0.870865	0.034956	0.018697
	L	48	25	23	0.998747		
	X_2 AN	55	16	39	0.869893	0.250368	0.209926
	DW	17	16	1	0.322760		
	X_3 TP	70	32	38	0.994694	0.024013	0.024141
	P	2	0	2	0.000000		
	X_4 TA	30	10	20	0.918299	0.026074	0.013604
	A	42	22	20	0.998364		
	X_5 TB	61	22	39	0.943226	0.124809	0.090263
	B	11	10	1	0.439497		

Berdasarkan Tabel 1 diperoleh hasil nilai *gain ratio* terbesar berada pada faktor umur dengan nilai *gain ratio* sebesar 0.209926. Hal ini berarti faktor umur akan menjadi *node* akar dari pohon algoritma C5.0 (*node* 1) sehingga cabang untuk *node* akar ada dua yaitu AN (*node* 2) dan DW (*node* 3)., kemudian dilanjutkan dengan penentuan cabang untuk masing-masing *node*. Selanjutnya dilakukan pembagian kelas pada cabang yang telah diperoleh dan proses tersebut diulang hingga setiap cabang memiliki kelas. Dalam penelitian ini analisis untuk penentuan cabang dilakukan hingga iterasi keempat. Karena tidak lagi memungkinkan untuk membuat cabang baru, maka proses pembuatan pohon dihentikan sehingga didapatkan sebuah pohon klasifikasi. Hasil akhir *decision tree* untuk metode algoritma C5.0 disajikan pada Gambar berikut:



Gambar 1 Pohon Klasifikasi Algoritma C5.0

Pada Gambar 1 terlihat bahwa variabel umur (X_2) yang menjadi akar (*node* 1) pada pohon algoritma C5.0. Hal ini dikarenakan variabel umur (X_2) memiliki nilai *gain ratio*

tertinggi dari variabel lainnya. Selanjutnya node 1 ini akan dijadikan sebagai pemisah untuk menentukan cabang pohon algoritma C5.0. Proses ini akan dilakukan secara terus menerus dan dihentikan apabila tidak ada lagi yang akan menjadi atribut serta memiliki kelasnya masing-masing. Dari Gambar 1 juga terlihat bahwa node 1 memiliki dua kategori yang dapat dilihat dari dua cabang yang ada pada gambar yaitu AN (node 2) dan DW (node 3) pada bagian kiri dan kanan. Pada node 2 menghasilkan variabel jenis kelamin (X_1) yang memiliki dua kategori yaitu P (node 4) dan L (node 5) sementara pada cabang bagian sebelah kanan pada kategori DW (node 3) menjadi terminal. Node 2 merupakan variabel jenis kelamin (X_1) yang memiliki dua kategori yang dapat dilihat dari dua cabang yang ada pada gambar yaitu kategori P (node 4) dan kategori L (node 5). Pada node 5 menghasilkan variabel riwayat pernah DBD (X_3) yang memiliki dua kategori yaitu TP (node 4) dan P (node 6) sementara pada cabang bagian sebelah kanan pada kategori P (node 7) menjadi terminal. Node 6 merupakan variabel riwayat penyakit penyerta (X_4) yang memiliki dua kategori yang dapat dilihat dari dua cabang yang ada pada gambar yaitu kategori TA (node 8) dan kategori A (node 9). Kemudian pada node 9 bercabang lagi hingga menghasilkan node yang terbagi menjadi dua cabang yaitu pada cabang bagian sebelah kiri memperlihatkan kategori TB (dengan hasil Ya DBD) sedangkan pada cabang bagian kanan memperlihatkan kategori B (dengan hasil TD DBD).

4. Simpulan dan Saran

Berdasarkan hasil analisis dan pembahasan klasifikasi dari data faktor-faktor penyebab terjadinya DBD dapat diperoleh kesimpulan sebagai berikut: berdasarkan analisis algoritma C5.0 terdapat lima variabel diantaranya variabel umur, jenis kelamin, riwayat pernah DBD, riwayat penyakit penyerta, serta status pekerjaan. Dari kelima variabel tersebut variabel umur merupakan atribut yang memberikan kontribusi terbesar yang dimana dapat disimpulkan bahwa dari kelima variabel tersebut dapat diperoleh bahwa faktor-faktor yang mempengaruhi terjadinya DBD secara signifikan di Kabupaten Bulungan adalah variabel umur, jenis kelamin, riwayat penyakit penyerta serta riwayat pernah DBD

Agar lingkup penelitian dapat diperluas, diharapkan agar ada pengembangan pada penelitian selanjutnya. Pengembangan metode klasifikasi juga dapat ditambahkan agar memperoleh hasil yang lebih akurat.

Daftar Pustaka

- Amalda, R. N., Millah, N., & Fitria, I. (2022). Implementasi Algoritma C5. 0 Dalam Menganalisa Kelayakan Penerima Keringanan Ukt Mahasiswa Itk. *Teorema: Teori Dan Riset Matematika*, 7(1), 101–116. <https://doi.org/http://dx.doi.org/10.25157/teorema.v7i1.6692>
- Chair, M., Nasution, Y. N., & Rizki, N. A. (2017). Aplikasi Klasifikasi Algoritma C4. 5 (Studi Kasus Masa Studi Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008). *Jurnal Informatika Mulawarman*, 12(1), 50–55. <https://doi.org/https://doi.org/10.31227/osf.io/85ajq>
- Ginting, S. L. B., Zarman, W., & Hamidah, I. (2014). Analisis Dan Penerapan Algoritma C4. 5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. *PROSIDING SNAST*, 263–272.
- Hikmah, M., & Kasmini H, O. W. (2015). Faktor yang berhubungan dengan kejadian kematian akibat demam berdarah dengue. *Unnes Journal of Public Health*, 4(4). <https://doi.org/https://doi.org/10.15294/ujph.v4i4>

- Latifah, K., Wibowo, S., & Nada, N. Q. (2018). Analisis dan Penerapan Algoritma C4. 5 Dalam Data Mining Untuk Menunjang Strategi Promosi Prodi Informatika UPGRIS. *J. Tek. Inform*, 11(2), 3–8. <https://doi.org/https://doi.org/10.15408/jti.v11i2.6706>
- Marwanty, M., & Wahyono, T. Y. M. (2019). Faktor lingkungan rumah dan kejadian demam berdarah dengue di kota palopo 2016. *Jurnal Epidemiologi Kesehatan Indonesia*, 2(1), 19–26. <https://doi.org/https://doi.org/10.7454/epidkes.v2i1.3106>
- Novrita, B., Mutahar, R., & Purnamasari, I. (2017). Analisis Faktor Risiko Kejadian Demam Berdarah Dengue Di Wilayah Kerja Puskesmas Celikah Kabupaten Ogan Komering Ilir. *Jurnal Ilmu Kesehatan Masyarakat*, 8(1).
- Prasetyo, E. (2014). *DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab*. ANDI.
- Pratiwi, R., Hayati, M. N., & Prangga, S. (2020). Perbandingan Klasifikasi Algoritma C5. 0 Dengan Classification and Regression Tree (Studi Kasus: Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 14(2), 267–278. <https://doi.org/https://doi.org/10.30598/barekengvol14iss2pp273-284>
- Rahmayanti, N. A., Azhar, Y., & Marthasari, G. I. (2020). Implementasi Algoritma C5. 0 Untuk Menganalisa Gejala Prioritas Pada Anak Yang Mengalami Bullying. *Jurnal Repositor*, 2(8). <https://doi.org/https://doi.org/10.22219/repositor.v2i8.30807>
- Respati, T., Raksanegara, A., Djuhaeni, H., Sofyan, A., Agustian, D., Faridah, L., & Sukandar, H. (2017). Berbagai faktor yang memengaruhi kejadian demam berdarah dengue di Kota Bandung. *ASPIRATOR-Journal of Vector-Borne Disease Studies*, 9(2), 91–96. <https://doi.org/http://dx.doi.org/10.22435/aspirator.v9i2.4509.91-96>
- Sulaiman, D., Arif, A., & Christyanti, R. D. (2021). Penerapan Algoritma Kruskal Pada Jaringan Kabel di Tanjung Selor Wahana Matematika dan Sains : Jurnal. *Wahana Matematika Dan Sains: Jurnal Matematika, Sains, Dan Pembelajarannya*, 15(2), 1–15.
- Syahdan, S., Arif, A., & Megawati, M. (2021). Analysis of the Factors that Cause Dengue Hemorrhagic Fever (DHF) Using Chi-Square Automatic Interaction Detection (CHAID). *International Journal of Natural Science and Engineering*, 5(3), 104–113. <https://doi.org/https://doi.org/10.23887/ijnse.v5i3.41123>
- Widyawati. (2021). *Data Kasus Terbaru DBD di Indonesia*. Kementerian Kesehatan. <https://sehatnegeriku.kemkes.go.id/baca/umum/20201203/2335899/data-kasus-terbaru-dbd-indonesia/>