

Pengelompokan Daerah Rawan Demam Berdarah Dengan Metode *K-Means Clustering*

Sri Handani Widiastuti^{1✉}, Rio Jumardi²

^{1,2}Sekolah Tinggi Teknologi Bontang

handaniwidiastuti@gmail.com

Abstract

The Dengue fever is a disease found in tropical and subtropical regions that is transmitted through the bite of the *Aedes aegypti* mosquito. Dengue fever is still a serious problem for public health in the city of Bontang. For that we need a system that can classify the areas of spread of dengue fever in the city of Bontang. The K-Means Clustering algorithm is a non-hierarchical data clustering algorithm that partitions data into one or more clusters or groups so that data with the same characteristics are grouped into the same cluster. To classify areas in the city of Bontang into three clusters, namely Sporadic, Potential, and Endemic. Three villages were selected as the centroid of each cluster, Bontang Kuala as the centroid of the sporadic cluster, Telihan as the centroid of the potential cluster, and Tanjung Laut as the centroid of the Endemic cluster. Village data, cluster data, and centroid data are processed using the K-Means Clustering algorithm, and application testing using Black Box Testing. The results of the application using the clustering method obtained information that there were five villages for sporadic clusters, five villages for potential clusters, and five villages for endemic clusters from a total of 15 villages used as samples for environmental health by the Bontang City Health Service.

Keywords: Mapping, Dengue Fever, K-Means, Potential, Endemic.

Abstrak

Demam Berdarah merupakan penyakit yang terdapat di wilayah tropis maupun subtropis yang ditularkan melalui gigitan nyamuk *Aedes Aegypti*. Demam berdarah masih merupakan salah satu masalah yang serius bagi kesehatan masyarakat di kota Bontang. Untuk itu diperlukan sebuah sistem yang dapat mengelompokkan daerah penyebaran demam berdarah di kota Bontang. Algoritma *K-Means Clustering* adalah salah satu algoritma data *clustering* non hierarki yang mempartisi data ke dalam satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama. Untuk mengelompokkan daerah di kota Bontang menjadi tiga *cluster* yaitu Sporadis, Potensial, dan Endemis. Dipilih tiga kelurahan sebagai *centroid* dari tiap-tiap *cluster*, Bontang kuala sebagai *centroid cluster* sporadis, Telihan sebagai *centroid cluster* potensial, dan Tanjung Laut sebagai *centroid cluster* Endemis. Data kelurahan, data *cluster*, serta data *centroid* diproses menggunakan algoritma *K-Means Clustering*, untuk pengujian aplikasi menggunakan *Black Box Testing*. Hasil aplikasi menggunakan metode *clustering* diperoleh informasi bahwa terdapat lima kelurahan untuk *cluster* sporadis, lima kelurahan untuk *cluster* potensial, dan lima kelurahan untuk *cluster* endemis dari total 15 kelurahan yang digunakan sebagai *sampel* untuk penyehatan lingkungan oleh Dinas Kesehatan Kota Bontang.

Kata kunci: Pengelompokan, Demam Berdarah, *K-Means*, Potensial, Endemis.

JIdT is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Demam berdarah merupakan salah satu jenis penyakit yang terdapat di wilayah tropis maupun subtropis [1], [2], [3]. Demam berdarah dapat menular karena gigitan nyamuk *Aedes Aegypti* [4]. Dampak dari demam berdarah dapat membuat suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai demam, mual/muntah, sakit kepala, nyeri perut, dan lekopenia [5].

Penyakit Demam Berdarah masih merupakan salah satu masalah yang serius bagi kesehatan masyarakat [6]. Dinas Kesehatan Kota Bontang mempunyai tugas untuk membantu masyarakat dalam hal penanggulangan kasus demam berdarah, karena tercatat sudah terjadi kurang lebih 548 kasus demam berdarah di Kota Bontang yang tersebar di 15

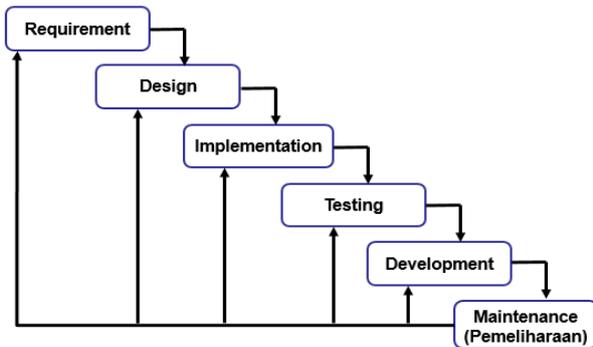
kelurahan. Untuk memudahkan Dinas Kesehatan Kota Bontang melakukan penanggulangan dan pencegahan kasus demam berdarah di Kota Bontang maka dilakukan pengelompokan berdasarkan jumlah penduduk dan jumlah kasus yang terjadi di 15 kelurahan di Kota Bontang. Agar Dinas Kesehatan Bontang dapat melakukan langkah yang tepat untuk menanggulangi dan mencegah terjadinya kasus demam berdarah ke depannya.

Pendataan jumlah penderita demam berdarah pada tiap kelurahan dicatat secara manual dan pengelompokannya juga secara konvensional. Sehingga yang ditandai hanya dua kelurahan dengan kasus tertinggi dengan warna merah tanpa pengkategorian apakah daerah tersebut masuk ke daerah rawan demam berdarah atau tidak. Oleh karena itu perlu suatu sistem *clustering* yang dapat

membantu dalam mengelompokkan daerah mana saja yang terindikasi rawan terserang demam berdarah berdasarkan atribut-atribut perhitungan yang sudah ditentukan. Metode yang dapat digunakan untuk membangun sistem *clustering* tersebut adalah algoritma yang dapat membantu dalam pengelompokan dari sejumlah data, yaitu algoritma *K-Means Clustering*. *K-Means Clustering* adalah salah satu algoritma data *clustering* non hierarki yang mempartisi data ke dalam satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama [7]. Dengan menggunakan algoritma tersebut maka data penderita demam berdarah pada tiap kelurahan yang masuk dalam daerah rawan demam berdarah diklasifikasikan masuk kelompok sporadis, endemis atau potensial..

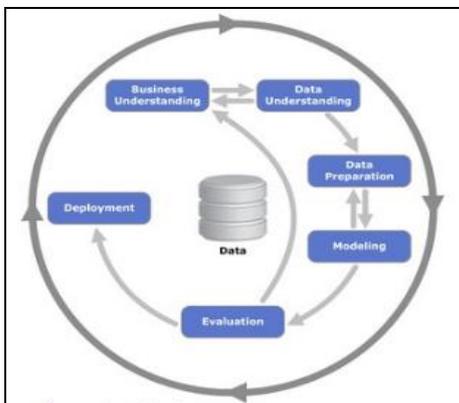
2. Metodologi Penelitian

Dalam membuat aplikasi metode pengembangan sistem SDLC (*System Development Life Cycle*) yang digunakan adalah model *waterfall* seperti yang terdapat pada Gambar 1.



Gambar 1. CRISP-DM Model

Proses pengelompokan (*clustering*) menggunakan model proses data mining Cross Industry Standard Process Model for Data Mining (CRISP-DM). Model ini terdiri dari 6 tahap yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*, yang disajikan pada Gambar 2.



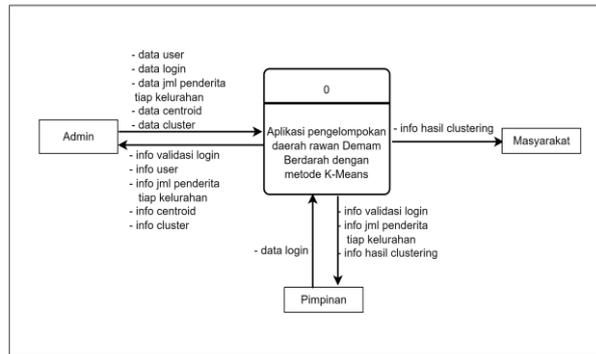
Gambar 2. CRISP-DM Model

2.1 Requirement

Tahap *requirement* ini dilakukan analisa kebutuhan dalam pembuatan atau pengembangan sistem antara lain melakukan observasi untuk pengambilan data yang dibutuhkan, spesifikasi *software* dan *hardware* yang dibutuhkan serta mengidentifikasi *user* yang akan menggunakan aplikasi.

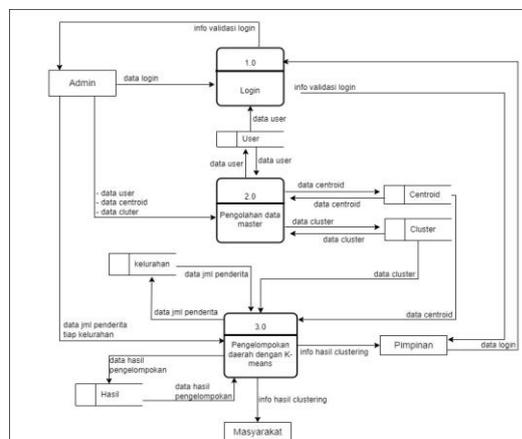
2.2 Perancangan Sistem

Pada tahapan ini dilakukan pembuatan gambaran rancangan aplikasi secara keseluruhan sebelum dilakukan coding, yang tersaji pada Gambar 3.



Gambar 3. Diagram Konteks Aplikasi Pengelompokan Daerah Rawan Demam Berdarah

Pada Gambar 3 rancangan aplikasi pengelompokan daerah rawan demam berdarah mempunyai tiga entitas yaitu admin yang memiliki akses menginputkan data user, login, jumlah penderita DB tiap kelurahan, centroid, cluster / kelas dan melihat atau mencetak hasil proses pengolahan tersebut. Entitas pimpinan memiliki akses login dan melihat informasi jumlah penderita DB pada tiap kelurahan dan hasil proses pengelompokan, kemudian entitas masyarakat hanya dapat melihat info pengelompokan daerah rawan demam berdarah yang dihasilkan oleh aplikasi. Adapun gambaran rancangan aplikasi secara rincinya adalah DFD level 0 aplikasi pengelompokan daerah rawan demam berdarah seperti yang terdapat pada Gambar 4.



Gambar 4. DFD Level 0 Aplikasi Pengelompokan Daerah Rawan Demam Berdarah

Pada DFD level 0 aplikasi pengelompokan daerah demam berdarah menggambarkan aktifitas yang dilakukan yaitu terdiri dari tiga proses utama, proses pengolahan data master, proses login dan proses pengelompokan daerah demam berdarah dengan *K-means*. Pada proses pengolahan data master admin menginputkan data user, cluster, centroid dan dapat melihat informasi data-data tersebut. Pada proses login admin dan pimpinan menginputkan data login dan melihat info validasi login. Pada proses pengelompokan daerah rawan demam berdarah admin menginputkan data jumlah penderita demam berdarah tiap kelurahan, pimpinan dan masyarakat dapat melihat informasi hasil pengelompokan daerah demam berdarah. Selanjutnya hasil clustering tersebut akan disajikan dalam bentuk tabel dan grafik yang dapat dicetak oleh Admin.

2.3 Metode K-Means Clustering

Clustering Analysis adalah merupakan salah satu metode analisis data utama, yang membantu mengidentifikasi pengelompokan objek data secara alami dari kumpulan data. Clustering adalah metode klasifikasi tanpa pengawasan (*unsupervised classification*) dan merupakan proses mempartisi satu set objek data menjadi satu set sub kelas atau kelompok penuh makna. Hal ini dapat dilakukan dengan menerapkan berbagai ukuran kesamaan dan jarak ke algoritma. *K-Means* yang digunakan untuk mengelompokkan data menjadi beberapa kelompok dengan beberapa *Cluster* [8]. Data – data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah *Cluster*. Dimana setiap *Cluster* memiliki titik pusat yang disebut *Centroid* [9]. Langkah-langkah dalam algoritma clustering K-Mean adalah sebagai berikut :

- Tentukan jumlah *cluster* (k) yang ingin dibentuk
- Tentukan titik pusat (*centroid*) awal secara random, dalam menentukan buah pusat cluster awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data *input*.
- Menghitung jarak terdekat setiap data dengan *centroid*. Untuk mengukur jarak antara data dengan pusat cluster digunakan *Euclidian Distance* dengan menggunakan Persamaan (1).

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (1)$$

Dimana *De* adalah jarak tiap data dengan centroids (*Euclidian distance*), *i* merupakan banyaknya data / objek, (*x,y*) adalah nilai/koordinat dari data /objek dan (*s,t*) adalah nilai/koordinat centroid.

- Tentukan posisi centroids yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada centroid yang sama. Pusat cluster

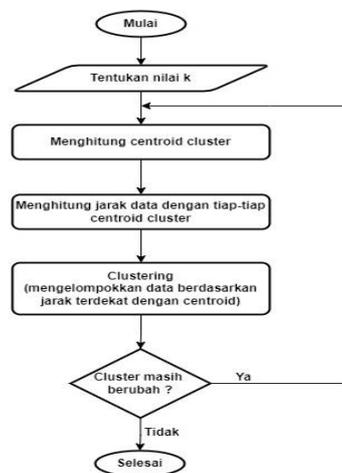
yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika hasil yang dicapai sudah konvergen (pusat cluster baru sama dengan pusat cluster lama). Perhitungan centroid yang baru menggunakan Persamaan (2).

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Dimana V_{ij} adalah centroid rata-rata ke *i*, N_i adalah jumlah anggota cluster ke-*i*, *k* merupakan indeks dari cluster dan *j* adalah indeks variabel serta X_{kj} adalah nilai data ke *k* variabel ke *j* untuk cluster tersebut.

- Kembali ke langkah 3 jika *cluster centroids* baru tidak sama dengan *cluster centroids* yang lama.

Gambaran langkah-langkah dalam algoritma K-Means terdapat pada Gambar 5.



Gambar 5. Flowchart Algoritma K-Means

3. Hasil dan Pembahasan

3.1 Perhitungan Algoritma K-Means

Berikut ini adalah data jumlah penduduk dan jumlah penderita DBD pada tiap kelurahan yang terdapat pada Tabel 1.

Tabel 1. Data Penderita DBD Pada Tiap Kelurahan (orang)

Kelurahan	Penduduk	Penderita
Api-Api	16.799	126
Gunung Elai	16.127	108
Bontang Baru	11.871	113
Bontang Kuala	4.801	17
Loktuan	20.973	214
Guntung	4.893	76
Tanjung Laut	16.141	196
Satimpo	8.195	55
Tj. Laut Indah	13.904	115
Berbas Pantai	9.968	82
Berbas Tengah	16.090	139
Bontang Lestari	4.629	101
Belimbing	12.993	73
Kanaan	4.080	63
Telihan	12.742	106

- a. Menentukan jumlah cluster/kelompok, pada penelitian ini ditentukan 3 cluster yaitu sporadis, potensial dan endemis
- b. Menentukan centroid awal pada tiap cluster secara random, cluster sporadic centroidnya adalah kelurahan Bontang Kuala, cluster potensial kelurahan Telihan dan cluster endemis kelurahan Tanjung Laut. Data centroid awal tersaji pada Tabel 2.

Tabel 2. Data Centroid (orang)

Centroid	Cluster	Penduduk	Penderita
Bontang Kuala	Sporadis	4.801	17
Telihan	Potensial	12.742	106
Tanjung Laut	Endemis	16.141	115

- c. Menghitung jarak tiap kelurahan terhadap masing-masing centroid dengan menggunakan persamaan (1). Berikut contoh yang dihitung berdasarkan cluster yaitu kelurahan Api-Api:

- a. Jarak Kelurahan Api-Api dengan titik pusat Cluster Sporadis yaitu ;

$$De = \sqrt{(jml\ penduduk\ Api - Api - jml\ penduduk\ Btg\ Kuala)^2 + (jml\ DBD\ Api - Api - jml\ DBD\ Btg\ Kuala)^2}$$

$$De = \sqrt{(16799 - 4801)^2 + (126 - 17)^2}$$

$$De = \sqrt{(11998)^2 + (109)^2}$$

$$De = \sqrt{(143952004) + (11881)}$$

$$De = \sqrt{(143963885)}$$

$$De = 11998,5$$

- b. Jarak Kelurahan Api-Api dengan titik pusat Cluster Potensial yaitu ;

$$De = \sqrt{(jml\ penduduk\ Api\ Api - jml\ penduduk\ Telihan)^2 + (jml\ DBD\ Api\ Api - jml\ DBD\ Telihan)^2}$$

$$De = \sqrt{(16799-12742)^2 + (126-106)^2}$$

$$De = \sqrt{(4057)^2 + (20)^2}$$

$$De = \sqrt{(16459249) + (400)}$$

$$De = \sqrt{(16459649)}$$

$$De = 4057,04$$

- c. Jarak Kelurahan Api-Api dengan titik pusat Cluster Endemis yaitu ;

$$De = \sqrt{(jml\ penduduk\ Api\ Api - jml\ penduduk\ TLaut)^2 + (jml\ DBD\ Api\ Api - jml\ DBD\ TLaut)^2}$$

$$De = \sqrt{(16799-16141)^2 + (126-115)^2}$$

$$De = \sqrt{(658)^2 + (11)^2}$$

$$De = \sqrt{(432964) + (121)}$$

$$De = \sqrt{(433085)}$$

$$De = 658,09$$

Dari perhitungan jarak kelurahan Api-Api terhadap tiga centroid, diperoleh hasil bahwa kelurahan Api-Api mempunyai jarak paling dekat dengan centroid Tanjung Laut. Maka kelurahan Api-Api masuk dalam

cluster endemis. Hasil perhitungan untuk seluruh kelurahan terdapat pada Tabel 3.

Tabel 3. Hasil Clustering Perhitungan Jarak

Cluster Sporadis		
Kelurahan	Penduduk	Penderita
Bontang Kuala	4.801	17
Guntung	4.893	76
Satimpo	8.195	55
Bontang Lestari	4.629	101
Kanaan	4.080	63
Rata-Rata	5.319	62
Cluster Potensial		
Bontang Baru	11.871	113
Tanjung Laut Indah	13.904	115
Berbas Pantai	9.968	82
Belimbing	12.993	73
Telihan	12.742	106
Rata-Rata	12.295	97
Cluster Endemis		
Api-Api	16.799	126
Gunung Elai	16.127	108
Loktuan	20.973	214
Tanjung Laut	16.141	196
Berbas Tengah	16.090	139
Rata-Rata	17.226	156

- d. Mengulang Langkah B dan melakukan iterasi perhitungan jarak dengan Ecludian Distance dengan centroid yang baru dan untuk pencarian nilai centroid baru dengan menggunakan persamaan (2). Berikut contoh perhitungan penentuan centroid baru.

Iterasi berhenti sampai hasil clustering sudah tidak berubah-ubah lagi.

3.2 Hasil Implementasi Algoritma K-Means dengan Program Aplikasi Komputer

- a. Halaman dashboard

Halaman dashboard user merupakan halaman awal Ketika user berhasil melakukan proses login. Tampilan halaman dashboard user dapat dilihat pada Gambar 6.



Gambar 6. Tampilan Halaman Dashboard User

- b. Halaman data kelurahan

Halaman ini menampilkan data kelurahan yang sudah diinputkan pada aplikasi computer. Tampilan halaman data kelurahan terdapat pada Gambar 7.

ID	Nama Kelurahan	Nilai
K5	Gurung	4833
K7	Tanjung Laut	16141
K8	Selimo	8195
K9	Tanjung Laut Indah	13904
K10	Beban Pantai	9968
K11	Beban Tengah	10990
K12	Bontang Laut	4629
K13	Bebanbing	12993
K14	Kamran	4100
K15	Telihan	12742

Gambar 7. Tampilan Halaman Data Kelurahan

d. Halaman data centroid

Halaman ini menampilkan data centroid awal yang telah diinputkan pada aplikasi computer. Tampilan halaman data centroid terdapat pada Gambar 8.

ID Centroid	Nama Kelurahan	Cluster	Nilai 1	Nilai 2	Edit	Hapus
C1	Bontang Kuala	Sporadis	4833	17		
C2	Telihan	Berpotensi	12742	106		
C3	Tanjung Laut	Endemis	16141	196		

Gambar 8. Tampilan Halaman Data Centroid

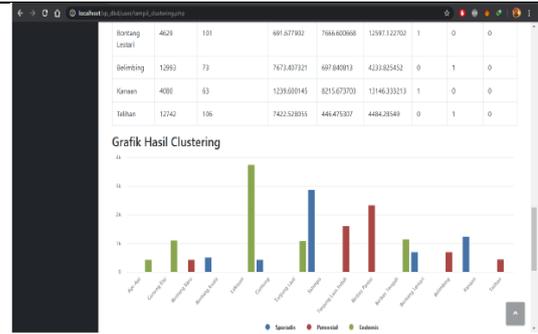
e. Halaman hasil clustering

Halaman ini menampilkan hasil perhitungan Euclidean Distance dan clustering. Tampilan halaman hasil clustering hasil perhitungan terdapat pada Gambar 9.

Nama Kelurahan	Jumlah Penduduk	Jumlah Penderita DBD (Dalam Setahun)	Jarak Ke Centroid 1	Jarak Ke Centroid 2	Jarak Ke Centroid 3	Sporadis	Potensial	Endemis
Api Api	1679	136	11479.579182	4333.488202	438.205036	0	0	1
Gurung Eli	16127	108	10887.4962	3831.412577	1188.07487	0	0	1
Bontang Baru	11671	113	6551.595403	424.671981	5355.577491	0	1	0
Bontang Kuala	4801	17	528.383442	7495.025544	12425.784207	1	0	0
Lukutan	20973	214	15654.13409	8678.177988	3747.439627	0	0	1
Gurung	4863	76	426.818729	7402.632066	12333.26337	1	0	0
Tanjung Laut	16141	196	10822.234075	3846.652663	1085.715138	0	0	1
Selimo	8195	55	2875.409522	4108.823356	9031.574489	1	0	0

Gambar 9. Tampilan Halaman Hasil Clustering

Tampilan halaman hasil clustering disertai grafik terdapat pada Gambar 10.



Gambar 10. Tampilan Halaman Hasil Clustering Dengan Grafik

4. Kesimpulan

Dari hasil penelitian implementasi Algoritma *K-Means Clustering* berhasil digunakan dalam mengelompokkan daerah rawan Demam Berdarah, dan kelurahan-kelurahan di kota Bontang dapat dibagi menjadi tiga *cluster*, yaitu *cluster* sporadis, *cluster* potensial, dan *cluster* endemis. Algoritma *K-Means Clustering* dapat digunakan sebagai metode untuk melakukan pengelompokan daerah rawan demam berdarah di kota Bontang. Hasil dari *clustering* menggunakan algoritma *K-Means clustering* dapat membantu Dinas Kesehatan kota Bontang dalam menjalankan program penyehatan lingkungan.

Daftar Rujukan

- Andriawan, F. R., Kardin, L., & HN, M. R. (2022). Hubungan Antara Status Gizi dengan Derajat Infeksi Dengue Pada Pasien Demam Berdarah Dengue. *Nursing Care and Health Technology Journal (NCHAT)*, 2(1), 8-15. DOI: <https://doi.org/10.56742/nchat.v2i1.33>
- AF, S. M., Wibowo, R. C. A., & Luthfin, A. (2022). Eksplorasi Sebaran Penyakit Demam Berdarah Dengue (DBD) dan Pneumonia di Kota Malang. *JUMANTIK (Jurnal Ilmiah Penelitian Kesehatan)*, 7(2), 134-140. DOI: <http://dx.doi.org/10.30829/jumantik.v7i2.10402>
- Tomia, A., Tuharea, R., & Hi, M. A. (2022). Kepadatan Larva Nyamuk Aedes aegypti berdasarkan House Index sebagai Indikator Surveilans Vektor Demam Berdarah Dengue di Kelurahan Tabona, Kecamatan Ternate Selatan. *JURNAL BIOSAINSTEK*, 4(2), 42-45. DOI: <https://doi.org/10.52046/biosainstek.v4i2.1064>
- Soedjadi, T. T. B., Tanjung, R., Syaputri, D., & Manalu, S. M. H. (2022). Container Dan Rumah Positif Jentik Nyamuk Aedes Aegypti Terhadap Kejadian Demam Berdarah Dengue. *Jurnal Ilmiah PANNMED (Pharmacist, Analyst, Nurse, Nutrition, Midwifery, Environment, Dentist)*, 17(1), 55-62. DOI: <https://doi.org/10.36911/panmed.v17i1.1263>
- Rahayu, E. P. F. (2022). Hipertermia (Demam Berdarah Dengue/DBD) pada Nona L Di Ruang Sambiloto Rumah Sakit Tk Ii Kartika Husada (Doctoral Dissertation, STIK Muhammadiyah Pontianak).
- Yuliana, R., Rahmaniati, M., Apriantini, I., & Triarjunet, R. (2022). Pemetaan Kerawanan dan Penentuan Prioritas Penanganan Penyakit Demam Berdarah Dengue di Kota Padang. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 5(5), 503-511. DOI: <https://doi.org/10.56338/mppki.v5i5.2278>
- Koteswara Rao, N. & Sridhar Reddy, G. (2012). Discovery of preliminary centroid using improved K-Means clustering

- Algorithm. *International Journal of Computer Science and Information Technologies*, 3(3), 4558-4561
- [8] Sihananto, A. N., Sari, A. P., Khariono, H., Fernanda, R. A., & Wijaya, D. C. M. (2022). Implementasi Metode K-Means Untuk Pengelompokan Kasus Covid-19 Tingkat Provinsi Di Indonesia. *Jurnal Informatika dan Sistem Informasi*, 3(1), 76-85. DOI: <https://doi.org/10.33005/jifosi.v3i1.472>
- [9] Putra, A. Z., Pinem, R. W., Silalahi, S., Gulo, F., & Liukhoto, J. A. A. (2022). Classification of Covid-19 Patient Spread Rate By Age and Region With K-Means Algorithm. *Sinkron: jurnal dan penelitian teknik informatika*, 7(3), 1085-1989. DOI: 10.33395/sinkron.v7i3.11603