

Pemetaan Kota/Kabupaten Endemis Demam Berdarah Dengue Dengan Analisis Data Science Menggunakan Algoritma Clustering

Mega Bagus Herlambang¹, Linda Theresia^{2*}, Salahuddin²

^{1,2} Jurusan Teknik Industri Institut Teknologi Indonesia

³ Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe

¹mega.herlambang@iti.ac.id · ^{2*}tarlind@yahoo.com (penulis korespondensi), ³salahuddintik@pnl.ac.id

Abstrak— Terdapat 102 nilai indeks entomologi daerah yang terdiri dari beberapa kota dan kabupaten di seluruh Indonesia. Dari 102 daerah tersebut perlu dilakukan pemetaan daerah menjadi beberapa kelompok (*cluster*). Data science sebagai keilmuan yang mengkaji pengolahan dan analisis data sangat cocok dalam melakukan pemetaan daerah dengan menggunakan algoritma *clustering*. Pada penelitian ini digunakan dua algoritma yaitu *K-Means clustering* dan *hierarchical clustering*. Dari hasil *clustering* didapatkan jumlah *cluster* yang terbaik adalah 3 kelompok *cluster*. Dasar penentuan dan evaluasi *cluster* dilakukan dengan melihat skor *silhouette* dan dendrogram. Hasil *cluster* menunjukkan terdapat 8 daerah yang masuk ke dalam kelompok yang memang angka persebaran jentik nyamuknya masih tinggi. Dengan demikian, maka diharapkan pemerintah dapat membuat skala prioritas pemberantasan jentik nyamuk demam berdarah pada delapan daerah yang masuk ke dalam *cluster* tersebut.

Kata kunci: Indeks entomologi, Demam berdarah, Klastering, K-means, Dendrogram

Abstract— There were 102 entomology-index values of dengue fever that comprised of cities and regencies in Indonesia. It was necessary to map these areas into a few clusters for the government to make a better policy for reducing the number of dengue fever in Indonesia. For this issue, data science as a study to process and analyse data was suitable to build clusters. To perform the study, two clustering algorithms were used: K-Means and hierarchical clustering. To evaluate the results, the silhouette coefficients values of a few numbers of K-clusters were compared and showed that the best number of clusters were three clusters. The last cluster consisted of eight cities and regencies with high-risk number of dengue fever. Therefore, the government should pay more attention to these regions and prioritize reducing the number of dengue fever spread in these areas

Keywords— Entomology index, Dengue fever, Clustering, K-means, Dendrogram

I. PENDAHULUAN

Saat ini perkembangan teknologi informasi dan komunikasi semakin maju. Ditambah dengan mudahnya akses internet membuat banyak informasi mudah didapatkan dan bisa diolah dengan kuantitas yang sangat besar yang sering disebut dengan istilah *big data*. Banyak perusahaan dan organisasi mulai berlomba-lomba untuk menjadi yang terbaik dan menjadi yang pertama dalam aspek teknologi pengolahan data. Tidak terbatas pada dunia industri, maka pemerintah maupun negara juga harus dapat memanfaatkan teknologi terbaru khususnya dalam hal pembuatan kebijakan.

Data Science sebagai sebuah keilmuan tentang pengolahan data sudah banyak juga digunakan baik di industri maupun pemerintahan untuk dapat mengolah data yang berukuran besar dan ditujukan untuk pengambilan keputusan. Keilmuan *data science* merupakan bagian dari keilmuan maju yang menjadi bagian dari revolusi industri 4.0. Dalam implementasinya *data science* memerlukan keilmuan lain di dalamnya seperti statistik, matematika, ilmu komputer, dan *machine learning* dan bisa digunakan untuk menyelesaikan masalah terkini baik di industri maupun pemerintahan.

Salah satu permasalahan yang muncul terkait dengan data adalah di bidang kesehatan. Tercatat bahwa nilai persebaran kasus demam berdarah terus meningkat [1]. Salah satu faktor yang menyebabkan meningkatnya angka demam berdarah adalah masih banyaknya genangan air yang menjadi sumber dari jentik nyamuk. Selain itu, masalah lain dari sisi pemerintah adalah belum adanya pemetaan daerah sehingga belum dapat dipetakan dengan baik mana daerah yang masih memerlukan upaya lebih tinggi dalam pemberantasan jentik nyamuk.

Oleh karena itu pemetaan daerah untuk menentukan daerah

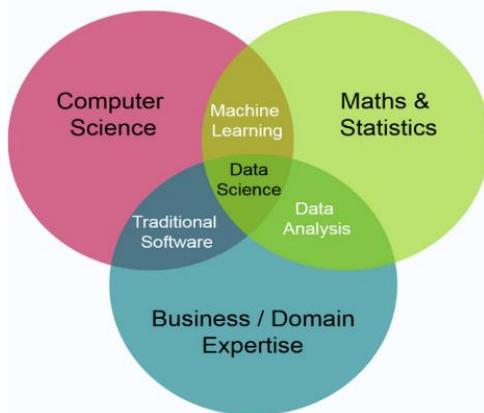
mana saja yang masuk ke dalam zona prioritas pemberantasan jentik nyamuk dirasa perlu untuk dilakukan. Dalam hal ini pendekatan *data science* adalah langkah yang tepat untuk bisa membuat pengelompokan daerah dalam bentuk kluster (*cluster*).

Penelitian ini bertujuan membuat pemetaan kota dan kabupaten di Indonesia menjadi beberapa kelompok (*cluster*) sehingga dapat didefinisikan wilayah mana saja yang mendapat prioritas utama pemberantasan jentik nyamuk. Untuk meningkatkan kemampuan manajemen informasi dari big data, penelitian clustering algorithm sangat diperhatikan pada saat ini [2]. Dalam metode tradisional, metode analisis big data clustering dari virtual resources utamanya adalah metode fuzzy C mean clustering, metode K mean clustering dan metode grid clustering [3]. Dengan menggunakan metode ini dapat dilakukan pemetaan kota dan kabupaten di Indonesia menjadi beberapa kelompok (*cluster*).

Penelitian ini bertujuan untuk memberikan usulan daerah hasil pemetaan yang layak mendapatkan prioritas tinggi untuk pemberantasan jentik nyamuk. Disamping itu penelitian ini berdampak di bidang kesehatan dan hasil pemetaan wilayah dapat digunakan oleh pemerintah khususnya bagi Kementerian Kesehatan untuk dapat menentukan prioritas daerah mana saja yang harus ditingkatkan pemberantasan jentik nyamuk. Dengan demikian, diharapkan tingkat persebaran kasus demam berdarah di Indonesia dapat diturunkan.

Penelitian ini menggunakan konsep data science. *Data science* merupakan keilmuan multidisiplin yang menggunakan kaidah-kaidah statistik, matematika, sains, dan bisnis untuk bisa mendapatkan *insights* (pemahaman mendalam) tentang data yang diolah yang nantinya dapat digunakan (untuk membantu) dalam pengambilan keputusan atau diaplikasikan di proses bisnis terkait. Seseorang yang berprofesi di bidang

data science disebut dengan *data scientist*. Untuk dapat menguasai *data science* dengan baik, maka setidaknya seorang *data scientist* harus menguasai beberapa keahlian. Drew Conway [4] menggambarkan sebuah diagram Venn yang menunjukkan keahlian apa saja yang harus dikuasai oleh mereka yang ingin terjun di *data science*.



Gambar 1. Diagram Venn keahlian *data scientist* (Conway, 2013)

Berdasarkan Gambar 1 di atas, salah satu keahlian yang harus dikuasai yang merupakan irisan antara ilmu komputer dengan matematika serta statistik adalah *machine learning*. Tidak hanya di keilmuan *data science*, dalam konteks keilmuan lain seperti kecerdasan buatan (*artificial intelligence*), maka *machine learning* ini merupakan komponen utama dalam proses analisis dan pemodelan data.

Machine learning atau dalam bahasa Indonesia disebut dengan pembelajaran mesin merupakan komponen utama dalam proses analisis data di *data science*. Tom Mitchell (1997) mendefinisikan *machine learning* sebagai sebuah proses pembelajaran mesin/komputer dari sekumpulan data untuk membuat sebuah model cerdas yang nantinya model ini bisa digunakan di dunia nyata dengan alat ukur (parameter performansi) yang sudah ditentukan untuk mengevaluasi baik tidaknya model tersebut saat digunakan [5].

Terdapat beberapa penerapan nyata dari *machine learning*. Contoh aplikasi dari *machine learning* di kehidupan sehari-hari seperti aplikasi email yang bisa mendeteksi mana email *spam* dan mana yang bukan. Kemudian *machine learning* juga bisa digunakan untuk memprediksi harga rumah, memprediksi curah hujan, memprediksi proyeksi ekonomi ke depan, dan masih banyak lagi. Secara umum, *machine learning* dibagi menjadi tiga jenis [6], antara lain: 1). *Supervised learning*; 2) *Unsupervised learning*; 3) *Reinforcement learning*. *Supervised learning* adalah cabang dari *machine learning* di mana kita sudah menentukan variabel mana yang akan diprediksi. Jika variabel yang diprediksi bertipe numerik (angka), maka menggunakan algoritma regresi. Kemudian jika variabel yang diprediksi bertipe kategori, maka menggunakan algoritma klasifikasi. *Unsupervised learning* merupakan bagian dari *machine learning* di mana tidak terdapat variabel yang ingin diprediksi. Tipe *machine learning* ini lebih fokuas mengelompokkan semua data yang dimiliki berdasarkan kemiripan fitur (*feature*) yang dimiliki oleh semua data yang ada. Beberapa algoritma *unsupervised learning* yang sering digunakan adalah *clustering*.

Reinforcement learning adalah tipe *machine learning* dengan menggunakan konsep *reward/punishment*, di mana model *machine learning* akan berusaha memaksimalkan

reward (atau dengan mengurangi *penalty/punishment*) jika model berhasil melakukan tujuan dari model yang dibuat. Penerapan *reinforcement learning* sering ditemui pada robotik, seperti misalnya mengajarkan bagaimana robot berjalan, berlari, dan melakukan aktivitas lainnya. Dalam konteks tujuan dari penelitian ini yaitu melakukan pemetaan, maka *unsupervised learning* adalah tipe *machine learning* yang tepat untuk mencapai tujuan tersebut. Dalam hal ini algoritma *clustering* dapat digunakan untuk memetakan data-data daerah, sehingga daerah-daerah yang memiliki banyak kesamaan akan membentuk *cluster* yang sama.

Pada penelitian ini digunakan *K-Means clustering*, merupakan tipe algoritma dari *unsupervised learning*. Algoritma ini berbasis pada kesamaan jarak (*distance*) antara data satu dengan yang lain dan membagi data ke dalam sejumlah kluster [7]. *K-means* merupakan salah satu algoritma pengelompokan yang dapat digunakan dalam berbagai keperluan, diantaranya untuk membuat peta inventory [8] dan mendeteksi tanah longsor [9,10,11,12]. Algoritma ini hanya dapat diimplementasikan pada data-data dengan atribut numerik. Berikut adalah langkah-langkah yang perlu dilakukan dalam algoritma *K-Means clustering*: 1). Menentukan jumlah *K* kluster; 2) Memilih secara random sebanyak *k* untuk menentukan pusat (*centroid*) dari masing-masing kluster; 3) Menghitung jarak *euclidean* dan masukan data ke *centroid* terdekat; 4) Menentukan *centroid* baru berdasarkan rata-rata data yang ada dari masing-masing kluster; 5) Kembali ke step 3 apabila masih dimungkinkan adanya perubahan nilai *centroid*. Jika tidak, selesai.

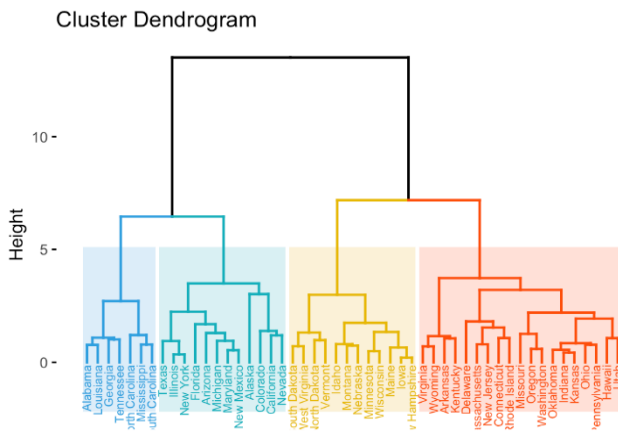
Formula untuk menghitung jarak *euclidean* dapat dilihat melalui rumus berikut:

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}$$

Di mana *D* adalah jarak dan a_k dan b_k adalah atribut ke-*k* dari objek data.

Disamping *K-Means clustering*, digunakan *hierarchical clustering*, merupakan bagian dari *unsupervised learning*. Algoritma ini menggunakan *distance matrix* sebagai kriteria untuk melakukan *clustering*. Berbeda dengan *K-Means clustering*, *hierarchical clustering* tidak membutuhkan penentuan jumlah kluster (*k*), namun membutuhkan *termination condition*. Pengelompokan dilakukan secara hierarkikal, di mana sekali langkah *merge* atau *split* dilakukan, tidak dapat mundur ke langkah sebelumnya. Sraphet (2022) pada penelitiannya menggunakan *hierarchical clustering* yang bertujuan memahami kekayaan database biologis dengan mempelajari evolusi pada tingkat molekuler terhadap area permukaan yang dapat diakses oleh pelarut invarian dan fitur struktural lipase *Amycolatopsis eburnea* untuk mengeksplorasi stabilitas dan evolusi enzim [13].

Ilustrasi pembentukan *clustering* melalui *hierarchical clustering* dapat dilihat melalui dendogram. *Clustering* dilakukan dengan memotong *dendogram* (pohon *cluster*) di level (batas) yang dikehendaki. Kemudian objek-objek yang terhubung dikelompokkan dalam satu *cluster*. Ilustrasi Dendogram ditunjukkan pada Gambar 2.



Gambar 2. Ilustrasi Dendrogram (Stackoverflow, 2022).

II. METODOLOGI PENELITIAN

Tahapan yang dilakukan dalam penelitian ini dijelaskan dibawah ini.

1) Pengumpulan data

Pengumpulan data, merupakan suatu studi yang dilakukan secara langsung dengan melakukan pengamatan pada objek data yang diteliti. Pada penelitian ini data yang digunakan adalah data-data yang terkait dengan jumlah penyebaran demam berdarah di propinsi di Indonesia.

2) Studi Pustaka

Studi Pustaka yang dilakukan di dalam penelitian ini menggunakan pendekatan metode CRISP-DM (The Cross Industry Standard Process for Data Mining) [14] dan beberapa algoritma *machine learning* dalam pembuatan modelnya. CRISP-DM adalah metodologi yang paling banyak digunakan untuk mengembangkan proyek Data Mining. Beberapa jurnal penelitian yang relevan juga menjadi bahan kajian dalam penelitian ini. Penelitian ini melakukan pemetaan daerah menggunakan metode clustering. Adapun tujuan pada penelitian ini adalah sebagai berikut: a) Mengetahui sebaran demam berdarah di 34 propinsi di Indonesia; b) Mengetahui pengelompokan daerah berdasarkan algoritma clustering; c) Membuat usulan skala prioritas penanganan demam berdarah berdasarkan hasil analisis clustering.

3) Pengolahan data

Untuk pengolahan data di dasarkan pada landasan teori yang dibahas dengan menggunakan metode yang telah ditentukan berdasarkan teori atau referensi yang ada. Data yang dibutuhkan ttd: a) Data indeks entomologi 34 propinsi; b) Data house index (HI) Jumlah rumah positif per rumah. Semakin tinggi HI, semakin tinggi risiko penularan DBD; c) Data container index (CI); d) Banyaknya jumlah penampungan air yang positif ditemukan jentik. Semakin tinggi CI, semakin tinggi risiko penularan DBD; e). Data breteau index (BI). Jumlah kontainer positif larva per rumah diperiksa. Semakin tinggi BI, semakin tinggi potensi penyebaran penyakit DBD; f) Data angka bebas jentik (ABJ). Persentase rumah atau tempat-tempat umum yang tidak ditemukan jentik. Semakin mendekati 100% semakin bebas dari jentik nyamuk. Sedangkan pengolahan data dilakukan dengan metode CRISP-DM, yang terdiri dari beberapa tahapan, antara lain: a) *Business understanding*; b) *Data understanding*; c) *Data preparation*; d) *Modeling*; e) *Evaluation*; f) *Deployment*.

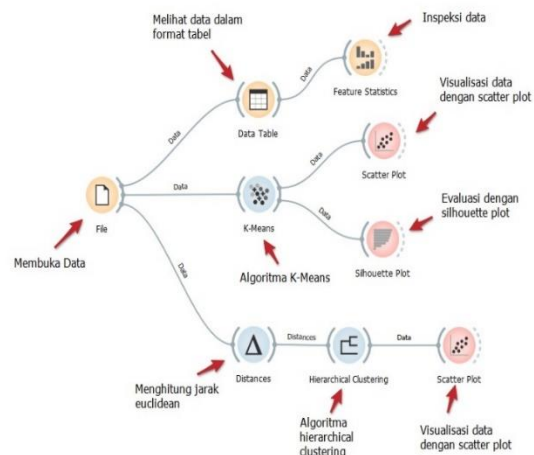
4) Analisis

Analisis dari hasil pengolahan data dan memberikan solusi untuk menyelesaikan permasalahan serta memberikan usulan kepada pihak terkait.

III. HASIL DAN PEMBAHASAN

1. Alur Proses di piranti lunak Orange

Dengan menggunakan *software* Orange, kita dapat membuat alur proses (*workflow*) mulai dari membuka data hingga didapatkan pembagian kelompok (*clusters*) menggunakan algoritma yang diinginkan. Tampilan dari *workflow* di *software* Orange secara keseluruhan dapat dilihat pada Gambar 3 di bawah ini:



Gambar 3. Tampilan *workflow* di Orange

Melalui gambar di atas dapat dilihat alur proses dari analisis data dari awal membuka data hingga tahapan evaluasi model *clustering* yang sebenarnya merupakan penjabaran dari alur proses CRISP-DM.

3.2. Business Understanding

Tahapan awal dari CRISP-DM adalah mendefinisikan masalah. Tahapan ini dijelaskan bahwa masalah yang dihadapi saat ini adalah belum adanya pemetaan daerah yang tepat dalam pemberantasan penyakit demam berdarah. Dengan adanya pemetaan, diharapkan pemerintah dapat memfokuskan anggaran kepada daerah-daerah yang memang tingkat penyakit demam berdarahnya masih tinggi. Oleh karena itu, diperlukan penggunaan algoritma *clustering* untuk membuat pemetaan daerah yang tepat.

3.3. Data Understanding

Tahapan kedua di metode CRISP-DM adalah memahami data yang diperlukan dan yang akan dianalisis. Dalam hal ini data didapatkan dari Kementerian Kesehatan dan proses analisisnya menggunakan *software* Orange.

Tahapan pertama pada Gambar 3 dimulai dengan membuka data menggunakan *widget* 'File'. Kemudian untuk melihat tampilan data dalam format tabel, digunakan *widget* 'Data Table'. Tampilan format tabel dari data yang dianalisis adalah seperti Gambar 4.



	Provinsi	Kab/Kota	HI (%)	CI (%)	BI (%)	ABJ (%)
1	Aceh	Kota Banda Aceh	52.67	27.33	78.00	47.33
2	Aceh	Kab. Aceh Besar	40.33	20.32	51.00	59.67
3	Aceh	Kota Lhokseum...	57.33	29.43	78.00	42.67
4	Sumatera Utara	Kab. Deli Serdang	31.50	15.74	40.39	68.50
5	Sumatera Utara	Kota Tebing Tin...	32.90	21.78	47.00	77.10
6	Sumatera Utara	Kota Pematang...	58.60	64.00	33.39	41.40
7	Sumatera Barat	Kota Bukit Tinggi	39.49	23.95	45.54	60.88
8	Sumatera Barat	Kota Padang	50.85	28.13	64.51	50.83
9	Sumatera Barat	Kab. Pesisir Sela...	46.00	32.42	51.00	57.05
10	Riau	Kota Pekanbaru	15.67	8.31	33.00	84.33
11	Riau	Kota Dumai	33.00	18.42	76.00	67.00
12	Riau	Kab. Bengkalis	31.00	18.14	90.00	69.00
13	Kepulauan Riau	Kota Tanjungpi...	49.70	14.70	71.70	50.30
14	Kepulauan Riau	Kab. Karimun	51.00	14.30	70.70	49.00

Gambar 4. Tampilan data yang diolah

Dapat diamati pada Gambar 4 bahwa data yang diolah merupakan data di setiap kota besar di setiap propinsi dengan total adalah sebanyak 102 baris data. Variabel yang digunakan sebagai dasar *clustering* terdiri dari 4 variabel utama, antara lain: a). Data *house index* (HI); b) Data *container index* (CI); c) Data *breteau index* (BI); d). Data angka bebas jentik (ABJ).

Pengamatan terhadap data di Orange dapat dilihat menggunakan *widget* 'Feature Statistics'. Hasil dari tampilan pengamatan data dapat dilihat pada Gambar 5.



Gambar 5. Pengamatan data menggunakan Feature Statistics

Melalui pengamatan di atas maka tidak terdapat nilai yang hilang (*missing values*). Kemudian semua data memiliki tipe data numerik. Dengan demikian, cocok untuk menggunakan model *clustering*.

3.4. Data Preparation

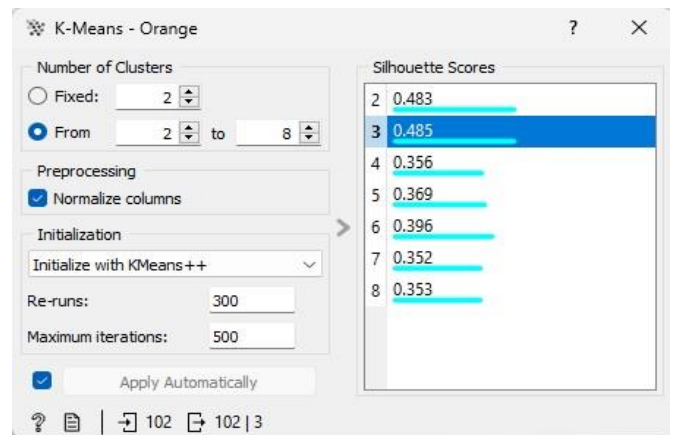
Sebelum data siap untuk dibuat modelnya, maka harus mengalami persiapan data. Sebagai contoh jika terdapat nilai kosong (*missing values*) maka harus dibuang nilai kosong atau dilakukan imputasi. Karena tidak terdapat nilai kosong, maka tidak perlu dilakukan imputasi. Selain itu, data juga sudah siap untuk dibuat model *clustering* karena semua variabel memiliki tipe numerik.

3.5. Modelling

Untuk membuat model dengan algoritma *clustering* maka dapat digunakan *widget* K-Means untuk membuat model K-Means *clustering*. Sementara untuk membuat model dengan algoritma *hierarchical clustering* dapat menggunakan *widget* Distances kemudian dilanjutkan dengan *widget* Hierarchical *clustering*. Visualisasi hasil *clustering* K-Means dan Hierarchical dapat dilihat menggunakan *widget* Scatter Plot.

3.6. Evaluasi

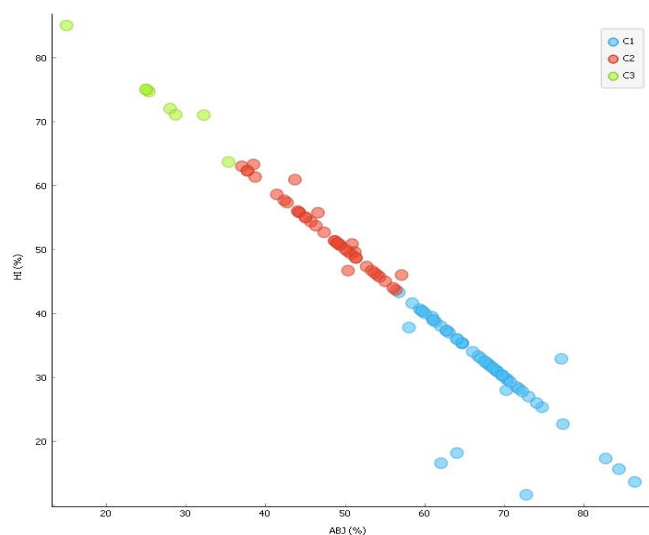
Proses evaluasi dari model *clustering* dapat dilakukan dengan menilai skor *silhouette coefficient*. Pada saat menggunakan *widget* K-Means kita bisa langsung melihat skor *silhouette* untuk beberapa nilai K. Ilustrasinya adalah seperti Gambar 6.



Gambar 6. Tampilan nilai silhouette di K-Means

Pada Gambar 6 di atas dapat dilihat skor nilai *silhouette* tertinggi adalah dengan nilai K=3 (0.485). Dengan demikian melalui algoritma K-Means maka jumlah *cluster* terbaiknya adalah sebanyak 3 buah.

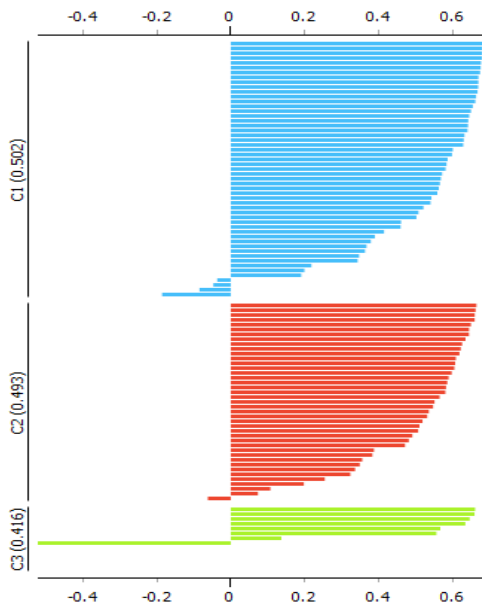
Tampilan hasil *cluster* dengan nilai K=3 melalui *widget* scatter plot dapat dilihat pada Gambar 7.



Gambar 7. Tampilan scatter plot untuk nilai K=3

Pada Gambar 7 di atas pembagian hasil *clustering*-nya sangat baik, di mana terdapat tiga *cluster* yang terbentuk dengan batas (pemisah) yang cukup jelas antara *cluster 1* (C1), *cluster 2* (C2), dan *cluster 3* (C3). Sumbu x yang digunakan dalam plot di atas adalah nilai Angka Bebas Jentik (ABJ), sementara sumbu y menggunakan nilai HI (*house index*).

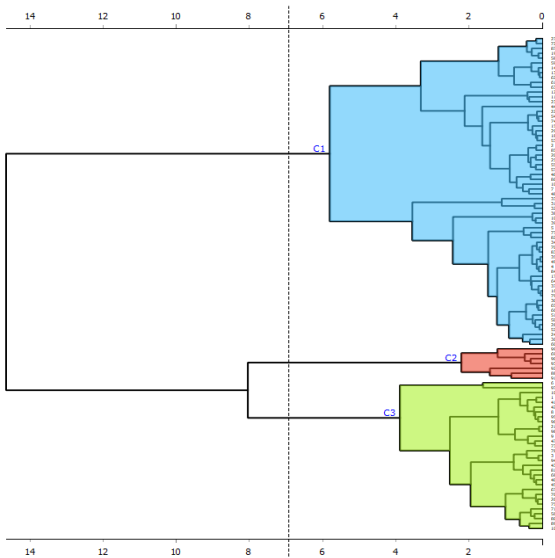
Penilaian hasil *clustering* dengan nilai K=3 juga bisa dilihat menggunakan *widget* silhouette plot. Hasil dari penggunaan *silhouette plot* adalah sebagai berikut:



Gambar 8. Tampilan silhouette plot untuk nilai K=3

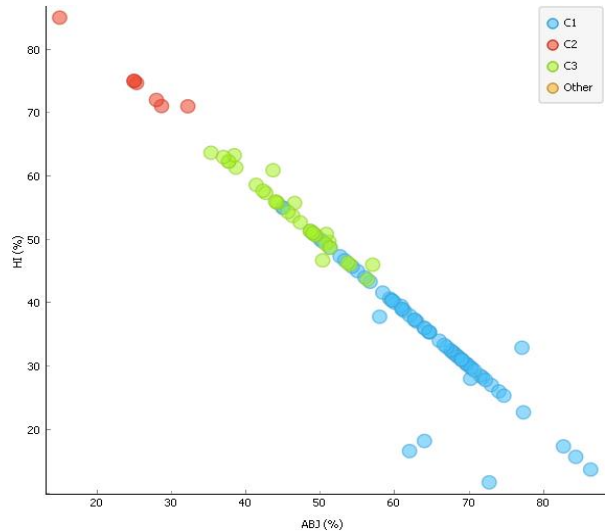
Pada Gambar 8 di atas dapat dilihat sebagian besar data berada di dalam pusat *clusternya* (pusat sentroid) yang ditunjukkan skor silhouette sebagian besar di atas 0. Hanya terdapat 1 nilai pada cluster 2 (warna merah) dan cluster 3 (warna hijau), 4 nilai pada cluster 1 (warna biru) yang sedikit keluar dari daerah *cluster*-nya. Hasil *silhouette plot* seperti ini dapat dikatakan cukup baik.

Proses evaluasi algoritma *hierarchical clustering* dilakukan dengan melihat dendrogram menggunakan *widget hierarchical plot*. Tampilan dari dendrogram yang dihasilkan adalah sebagai berikut:



Gambar 9. Tampilan dendrogram yang dihasilkan

Tampilan dendrogram pada Gambar 9 menunjukkan bahwa pembagian *cluster* yang terbaik adalah menjadi 3 *cluster*, di mana hasil dari *hierarchical clustering* sama dengan *K-Means clustering*. Hasil *clustering* dengan *hierarchical clustering* dapat dilihat dengan scatter plot. Ilustrasinya adalah seperti pada Gambar 10.



Gambar 10. Tampilan scatter plot dari hierarchical clustering

Scatter plot yang dihasilkan dari *hierarchical clustering* menunjukkan pembagian *cluster* yang cukup baik, walaupun terdapat sedikit pencampuran pada *cluster* 2 dan 3, namun masih didapat perbedaan yang baik. Dengan demikian pemilihan jumlah *cluster* menjadi 3 kelompok *cluster* adalah pembagian yang terbaik.

3.7. Deployment

Proses *deployment* yang paling sederhana adalah dengan menuliskan laporan (*report*) dari keseluruhan proses modeling dari awal hingga akhir. Dalam hal ini *deployment* dituangkan dalam bentuk laporan penelitian.

3.8. Hasil Cluster

Dengan menggunakan algoritma *K-Means* dan *Hierarchical Clustering*, didapatkan usulan jumlah *cluster* yang terbaik adalah 3 buah. Tampilan data pembentukan *cluster* dalam Tabel I.

Dari hasil pembagian data di Tabel I dan membandingkannya dengan scatter plot pada Gambar 7, maka bisa dilihat penggolongan setiap kelompok *cluster* sebagai berikut:

- *Cluster* 1 (C1) merupakan daerah-daerah yang sudah baik, dilihat dari tingkat angka bebas jentik yang sudah baik kemudian skor *house index* yang rendah (semakin rendah skor HI maka semakin rendah tingkat penyebaran jentik nyamuk). Daerah-daerah yang masuk ke kelompok ini merupakan kota-kota besar di Indonesia seperti beberapa kota di propinsi

TABEL I
DATA WILAYAH PEMBAGIAN *CLUSTER*

No	HI (%)	CI (%)	BI (%)	ABJ (%)	Cluster	Provinsi	Kab/Kota
1	40.33	20.32	51	59.67	C1	Aceh	Kab. Aceh Besar
2	31.5	15.74	40.39	68.5	C1	Sumatera Utara	Kab. Deli Serdang
3	32.9	21.78	47	77.1	C1	Sumatera Utara	Kota Tebing Tinggi
4	39.49	23.95	45.54	60.88	C1	Sumatera Barat	Kota Bukit Tinggi
5	15.67	8.31	33	84.33	C1	Riau	Kota Pekanbaru
6	33	18.42	76	67	C1	Riau	Kota Dumai
7	31	18.14	90	69	C1	Riau	Kab. Bengkalis
8	40.4	11.9	49.4	59.6	C1	Kepulauan Riau	Kota Batam
9	32.5	12.4	45.4	67.5	C1	Bengkulu	Kab. Bengkulu Tengah
10	35.4	14.1	41.3	64.6	C1	Bengkulu	Kab. Rejang Lebong
11	37.3	14.4	48	62.7	C1	Bengkulu	Kota Bengkulu
12	40	17.41	53.33	60	C1	Jambi	Kab. Batanghari
13	36	12.8	48	64	C1	Sumatera Selatan	Kota Lubuk Linggau
14	30.3	22.9	78.3	69.7	C1	Sumatera Selatan	Kota Palembang
15	22.7	14.9	52.3	77.3	C1	Sumatera Selatan	Kota Prabumulih
16	41.6	15.8	56.7	58.4	C1	Bangka Belitung	Kab. Bangka Tengah
17	29.3	10.3	39.6	70.7	C1	Lampung	Kab. Pringsewu
18	39	15.4	46	61	C1	Lampung	Kota Bandar Lampung
19	27.8	9.2	31.1	72.2	C1	Lampung	Kota Metro
20	16.59	38	48.33	62	C1	Banten	Kota Cilegon
21	18.18	36	52.67	64	C1	Banten	Kota Serang
22	11.65	27.3	32.33	72.7	C1	Banten	Kota Tangerang Selatan
23	29.67	14.56	45.67	70.33	C1	DKI Jakarta	Kota Jakarta Barat
24	31	14.38	39.33	69	C1	DKI Jakarta	Kota Jakarta Selatan
25	25.33	13.87	39.33	74.67	C1	DKI Jakarta	Kota Jakarta Timur
26	33.33	11.92	47	66.67	C1	Jawa Barat	Kota Bandung
27	17.33	9.97	37.67	82.67	C1	Jawa Barat	Kota Sukabumi
28	13.67	9.93	31.33	86.33	C1	Jawa Barat	Kota Tasikmalaya
29	43.3	16.9	3.2	56.7	C1	DI Yogyakarta	Kab. Bantul
30	35.33	21.21	37.7	64.67	C1	Jawa timur	Kab. Malang
31	35.33	26.17	53.3	64.67	C1	Jawa timur	Kab. Ngawi
32	31.5	14.7	38	68.5	C1	Sulawesi Utara	Kab. Minahasa
33	29.8	12	42	70.2	C1	Sulawesi Utara	Kab. Minahasa Utara
34	28	11.5	31.7	70.2	C1	Sulawesi Utara	Kota Manado
35	28.5	9.7	38.19	71.5	C1	Sulawesi Barat	Kab. Polewali Mandar
36	38.7	13.1	50	61.3	C1	Sulawesi Barat	Kab. Mamuju Tengah
37	37.3	11.6	52.33	62.7	C1	Sulawesi Barat	Kab. Mamuju Utara
38	37	16.4	58.7	63	C1	Sulawesi Selatan	Kota Palopo
39	39	17.9	60.3	61	C1	Sulawesi Selatan	Kota Makassar
40	28.2	17.6	34.88	71.8	C1	Sulawesi Tengah	Kota Palu
41	34	10	39	66	C1	Gorontalo	Kota Gorontalo
42	27	9	35	73	C1	Gorontalo	Kab. Gorontalo
43	26	8	33	74	C1	Gorontalo	Kab. Bone Bolango
44	32	14.9	45.3	68	C1	Kalimantan Tengah	Kota Palangkaraya
45	36	11.34	50	64	C1	Kalimantan Selatan	Kab. Hulu Sungai Utara
46	32.33	21.75	49	67.67	C1	Kalimantan Timur	Kab. Kutai Kertanegara
47	32	12.2	44.7	68	C1	Kalimantan Utara	Kab. Bulungan
48	30.33	22.83	37.67	69.67	C1	Bali	Kab. Badung
49	31	14.38	44.67	69	C1	Bali	Kab. Gianyar
50	30.33	16.31	39.33	69.67	C1	Bali	Kota Denpasar
51	40.67	19.73	57.67	59.33	C1	Nusa Tenggara Barat	Kota Mataram
52	38	18.28	45	62	C1	Nusa Tenggara Barat	Kab. Lombok Tengah
53	37.79	22.89	64.7	57.98	C1	Papua	Kab. Merauke
54	52.67	27.33	78	47.33	C2	Aceh	Kota Banda Aceh
55	57.33	29.43	78	42.67	C2	Aceh	Kota Lhokseumawe
56	58.6	64	33.39	41.4	C2	Sumatera Utara	Kota Pematangsiantar
57	50.85	28.13	64.51	50.83	C2	Sumatera Barat	Kota Padang

No	HI (%)	CI (%)	BI (%)	ABJ (%)	Cluster	Provinsi	Kab/Kota
58	46	32.42	51	57.05	C2	Sumatera Barat	Kab. Pesisir Selatan
59	49.7	14.7	71.7	50.3	C2	Kepulauan Riau	Kota Tanjungpinang
60	51	14.3	70.7	49	C2	Kepulauan Riau	Kab. Karimun
61	44	18.96	61	56	C2	Jambi	Kota Jambi
62	48.67	35.76	73.67	51.33	C2	Jambi	Kab. Muaro Jambi
63	55.8	23.5	87.4	44.2	C2	Bangka Belitung	Kab. Bangka Barat
64	46.7	16.4	64.7	53.3	C2	Bangka Belitung	Kota Pangkal Pinang
65	61.33	28.77	84	38.67	C2	Jawa tengah	Kab. Blora
66	51.33	29.33	78.33	48.67	C2	Jawa tengah	Kab. Brebes
67	51	26.19	70.33	49	C2	Jawa tengah	Kota Tegal
68	62.3	30.3	64	37.7	C2	DI Yogyakarta	Kab. Sleman
69	62.3	30.3	79.3	37.7	C2	DI Yogyakarta	Kab. Gunung Kidul
70	46.33	31.49	62.7	53.67	C2	Jawa timur	Kab. Blitar
71	53.7	25.4	92.7	46.3	C2	Sulawesi Selatan	Kab. Bone
72	45	17.9	62.33	55	C2	Sulawesi Tengah	Kab. Morowali Utara
73	48.7	11.8	66	51.3	C2	Sulawesi Tengah	Kab. Tolitoli
74	55	15	75	45	C2	Sulawesi Tenggara	Kota Kendari
75	50	15	69	50	C2	Sulawesi Tenggara	Kab. Bombana
76	55	17	87	45	C2	Sulawesi Tenggara	Kab. Kolaka
77	60.91	23.39	99.09	43.66	C2	Kalimantan Barat	Kab. Mempawah
78	63.28	27.8	75.74	38.46	C2	Kalimantan Barat	Kab. Kuburaya
79	49.3	23.8	98.3	50.7	C2	Kalimantan Tengah	Kab. Kotawaringin Barat
80	45.7	15.3	70	54.3	C2	Kalimantan Tengah	Kab. Kotawaringin Timur
81	43.67	31.58	66	56.33	C2	Kalimantan Selatan	Kab. Hulu Sungai Selatan
82	54.33	22.74	85.33	45.67	C2	Kalimantan Selatan	Kab. Tabalong
83	56	23.03	80.67	44	C2	Kalimantan Timur	Kota Samarinda
84	57.67	30.75	92.67	42.33	C2	Kalimantan Timur	Kota Balikpapan
85	50.7	25.8	92.7	49.3	C2	Kalimantan Utara	Kab. Nunukan
86	63	25.4	83	37	C2	Kalimantan Utara	Kota Tarakan
87	47.33	20.62	60	52.67	C2	Nusa Tenggara Barat	Kab. Lombok Timur
88	46	22.91	79.33	54	C2	Nusa Tenggara Timur	Kab. Kupang
89	55.81	29.21	78.14	44.19	C2	Maluku Utara	Kab. Halmahera Timur
90	51.33	29.64	70.97	48.67	C2	Maluku Utara	Kota Ternate
91	50.67	28.47	68.09	49.33	C2	Maluku Utara	Kab. Tidore Kepulauan
92	49.61	35.55	70.54	51.18	C2	Papua Barat	Kab. Sorong
93	55.74	30.74	56.7	46.56	C2	Papua	Kota Jayapura
94	46.69	26.38	87.4	50.33	C2	Papua	Kab. Jayapura
95	71.04	31.97	158.2	28.69	C3	Kalimantan Barat	Kab. Ketapang
96	75	46.52	147.33	25	C3	Nusa Tenggara Timur	Kab. Timor Tengah Selatan
97	75	36.31	146.33	25	C3	Nusa Tenggara Timur	Kota Kupang
98	72	58.28	138	28	C3	Maluku	Kab. Kepulauan Aru
99	85	49.19	181.67	15	C3	Maluku	Kab. Maluku Tenggara
100	74.67	36.68	159.67	25.33	C3	Maluku	Kota Ambon
101	63.67	48.35	83	35.33	C3	Papua Barat	Kota Sorong
102	71	33.66	115	32.22	C3	Papua Barat	Kab. Manokwari

DKI Jakarta, kota Tangerang Selatan, kota Bandung, dan sebagainya. Daerah ini bisa dikategorikan sebagai daerah yang harus dipertahankan tingkat pemberantasan jentik nyamuknya.

- *Cluster (C2)* merupakan daerah yang cukup baik, dilihat dari tingkat nilai angka bebas jentik yang cukup baik (namun masih di bawah C1) dan skor HI yang sedang. Daerah ini bisa dikategorikan sebagai daerah yang perlu ditingkatkan upaya pemberantasan jentik nyamuk.
 - *Cluster (C3)* merupakan daerah yang masih butuh

banyak perbaikan. Kondisi ini dilihat dari skor angka bebas jentik yang masih rendah dan skor HI yang paling tinggi. Dengan demikian, daerah ini bisa dikategorikan sebagai daerah prioritas penanganan pemberantasan jentik nyamuk bagi pemerintah. Delapan daerah yang masuk ke dalam *cluster* ini antara lain: Kabupaten Ketapang, Kabupaten Timor Tengah Selatan, Kota Kupang, Kabupaten Kepulauan Aru, Kabupaten Maluku Tenggara, Kota Ambon, Kota Sorong, dan Kabupaten Manokwari.

IV. KESIMPULAN

Temuan penelitian menunjukkan bahwa keilmuan *data science* dapat digunakan untuk membantu proses analisis lebih lanjut terkait data yang dimiliki, sehingga dapat digunakan dalam proses pengambilan keputusan atau pembuatan kebijakan.

Algoritma *clustering* dapat digunakan untuk membentuk *cluster* dari data indeks entomologi jentik nyamuk. Algoritma *clustering* yang digunakan dalam penelitian ini adalah algoritma *K-Means clustering* dan *hierarchical clustering*. Melalui proses evaluasi dari jumlah *cluster* yang terbentuk, maka dapat disimpulkan jumlah *cluster* yang terbaik adalah sebanyak 3 *cluster*. Dari tiga *cluster* yang terbentuk, maka pemerintah dapat memprioritaskan untuk fokus meningkatkan pemberantasan jentik nyamuk di daerah-daerah yang masuk di *cluster* ketiga. Delapan daerah yang menjadi prioritas antara lain: Kabupaten Ketapang, Kabupaten Timor Tengah Selatan, Kota Kupang, Kabupaten Kepulauan Aru, Kabupaten Maluku Tenggara, Kota Ambon, Kota Sorong, dan Kabupaten Manokwari. Temuan penelitian sejalan dengan filosofi Machine Learning [15], [16] yang menyatakan bahwa sistem memiliki: (i) tugas yang harus dilakukan, (ii) kinerja untuk mengoptimalkan parameter sistem. Hal ini ditunjukkan oleh temuan penelitian yang dapat mengoptimalkan cara pemberantasan jentik nyamuk dengan memprioritaskan pada daerah yang masuk di *cluster* ketiga.

Penelitian ini dapat menjadi masukan bagi pemberantasan jentik nyamuk kepada Kementerian Kesehatan Republik Indonesia untuk dapat memprioritaskan delapan daerah yang masuk di *cluster* 3 sebagai daerah yang wajib ditingkatkan upaya pemberantasan jentik nyamuknya. Dengan demikian, diharapkan semakin sedikit angka penyebaran jentik nyamuk di Indonesia, khususnya di delapan daerah dalam *cluster* tersebut.

REFERENSI

- [1] Sehatnegeriku.(2022). <https://sehatnegeriku.kemkes.go.id/baca/umum/20220615/0240172/kasus-dbd-meningkat-kemenkes-galakkan-gerakan-1-rumah-1-jumantik-g1r1j/>
- [2] DING J, LI A X, HU Z Q, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks [C] //Proceedings of the 2017 International Conference on Medical Computing and Computer-Assisted Intervention. Berlin: Springer, 2017: 559-567
- [3] KAMNITSAS, K, LEDIG C, NEWCOMBE V F, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Medical Image Analysis, 2016, 36: 61-78.
- [4] Conway, D. (2013). <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- [5] Tom M. Mitchell, Machine Learning, McGraw Hill, 1997
- [6] Herlambang, M. B. (2022). Data Science A-Z Tanpa Coding.
- [7] Larose, D. T., Larose, C. D. (2015). Data Mining and Predictive Analytics. Wiley Publication.
- [8] Barbu M, Radoi A, Suci G. Landslide monitoring using convolutional autoencoders. Proceedings of 12th International Conference on Electrical Computers Artif Intelligence, ECAI 2020. Vol. 826452; 2020. p. 1-6
- [9] Keyport RN, Oommen T, Martha TR, Sajinkumar KS, Gierke JS. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. Int J Appl Earth Obs Geoinf. 2018;64(Febuary):1-11
- [10] Tran CJ, Mora OE, Fayne JV, Gabriela Lenzano M. Unsupervised classification for landslide detection from airborne laser scanning. Geosci (Switz). 2019;9(5):221.
- [11] Tehrani FS, Santinelli G, Herrera Herrera M. Multi-regional landslide detection using combined unsupervised and supervised machine learning. Geomatics Nat Hazards Risk. 2021;12(1):1015-38. doi: 10.1080/19475705.2021.1912196.
- [12] Gorsevski PV, Jankowski P, Gessler PE. Spatial prediction of landslide hazard using fuzzy k-means and Dempster-Shafer theory. Trans GIS. 2005;9(4):455-74
- [13] Sraphet, S.; Javadi, B. Application of Hierarchical Clustering to Analyze Solvent-Accessible Surface Area Patterns in Amycolatopsis lipases. Biology 2022, 11, 652. <https://doi.org/10.3390/biology11050652>
- [14] O. Marbán, G. Mariscal and J. Segovia, Data Mining and Knowledge Discovery in Real Life Applications, 2009.
- [15] David Westgarth, Coronavirus pandemic: A time for reflection, BDJ In Practice 33, 4-4, 4 May 2020.
- [16] T. Tanaka; T. M. Mitchell, Embedding learning in a general framebased architecture, IEEE International Workshop on Tools for Artificial Intelligence, Year: 1989 Pages: 77 - 84.