



**ANALISIS *CLUSTER* UNTUK MENENTUKAN WILAYAH  
BERPOTENSI PENYAKIT DEMAM BERDARAH *DENGUE*  
DI JAWA BARAT**

**SKRIPSI**

**MAULIDA AFIFAH  
2010512066**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI S1 SISTEM INFORMASI  
2024**



**ANALISIS *CLUSTER* UNTUK MENENTUKAN WILAYAH  
BERPOTENSI PENYAKIT DEMAM BERDARAH *DENGUE*  
DI JAWA BARAT**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Komputer**

**MAULIDA AFIFAH  
2010512066**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI S1 SISTEM INFORMASI  
2024**

# PERNYATAAN ORISINALITAS

## PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Maulida Afifah

NIM : 2010512066

Program Studi : S1-Sistem Informasi

Judul : Analisis *Cluster* Untuk Menentukan Wilayah Berpotensi Penyakit  
Demam Berdarah *Dengue* Di Jawa Barat

Tanggal : 05 Juli 2024

Bilamana di kemudian ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses dengan ketentuan yang berlaku.

Jakarta, 05 Juli 2024

Yang Menyatakan,



(Maulida Afifah)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI  
UNTUK KEPENTINGAN AKADEMIS**

---

Sebagai civitas akademik Universitas Pembangunan Nasional "Veteran"  
Jakarta, saya yang menandatangani pernyataan ini:

Nama : Maulida Afifah

NIM : 2010512066

Program Studi : S1 Sistem Informasi

Fakultas : Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (Non-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul:

**ANALISIS *CLUSTER* UNTUK MENENTUKAN WILAYAH  
BERPOTENSI PENYAKIT DEMAM BERDARAH *DENGUE* DI JAWA  
BARAT**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/ formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta  
Pada Tanggal : 05 Juli 2024

Yang Menyatakan,



(Maulida Afifah)

# LEMBAR PENGESAHAN

## LEMBAR PENGESAHAN

Skripsi ini diajukan oleh

Nama : Maulida Afifah  
NIM : 2010512066  
Program Studi : S1 – Sistem Informasi  
Judul : Analisis *Cluster* Untuk Menentukan Wilayah Berpotensi Penyakit Demam Berdarah *Dengue* Di Jawa Barat

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.  
Dosen Pembimbing 1



Musthofa Galih Pradana, S.Kom., M.Kom.  
Dosen Pembimbing 2



I Wayan Widi, S.Kom., MTI.  
Dosen Penguji 1



Mohamad Bayu Wibisono, S.Kom., MM.  
Dosen Penguji 2



Prof. Dr. H. Supriyanto, ST., M.Sc., IPM  
Dekan Fakultas Ilmu Komputer



Anita Muliawati, S.Kom., MTI.  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 02 Juli 2024

# **ANALISIS *CLUSTER* UNTUK MENENTUKAN WILAYAH BERPOTENSI PENYAKIT DEMAM BERDARAH *DENGUE* DI JAWA BARAT**

**Maulida Afifah**

## **ABSTRAK**

Demam Berdarah *Dengue* (DBD) adalah salah satu masalah kesehatan masyarakat yang serius di Indonesia, khususnya di Jawa Barat, wilayah dengan kasus Demam Berdarah *Dengue* tertinggi pada tahun 2022. Dengan latar belakang peningkatan kasus DBD yang berhubungan dengan faktor-faktor seperti kepadatan penduduk dan curah hujan yang tinggi, penelitian ini mengaplikasikan metode *Clustering* K-Means untuk mengidentifikasi kabupaten/kota di Jawa Barat berdasarkan potensi DBD. Menggunakan data dari tahun 2020 hingga 2022 yang mencakup kepadatan penduduk, jumlah kasus DBD, dan kejadian banjir yang didapat dari website Pemerintah Jawa Barat. Hasil analisis menunjukkan bahwa nilai K optimal adalah 4, yang didukung oleh grafik metode elbow yang menunjukkan penurunan nilai Sum of Squared Errors (SSE) yang melambat setelah jumlah cluster mencapai 4. Penelitian ini berhasil mengelompokkan wilayah tersebut ke dalam empat cluster: Endemis Tinggi merupakan wilayah dengan jumlah kasus dan kepadatan penduduk tinggi, Endemis Sedang merupakan wilayah dengan jumlah kasus dan jumlah banjir tinggi, Potensial merupakan wilayah dengan kasus DBD jarang terjadi dan kepadatan penduduk serta frekuensi banjir yang rendah. Terakhir, dan Sporadis yang menunjukkan jumlah kasus DBD yang paling rendah, kepadatan penduduk, dan jumlah banjir yang juga paling rendah. Hasil visualisasi ditampilkan dalam bentuk peta interaktif yang memudahkan pemahaman distribusi *cluster* di berbagai wilayah tahun 2020-2022. Hasil ini bermanfaat untuk mengarahkan intervensi dan sumber daya kesehatan secara lebih efektif di Jawa Barat, serta mendukung pengambilan keputusan berbasis data dalam pengendalian DBD.

**Kata kunci:** Demam Berdarah *Dengue*, *Clustering* K-Means, visualisasi data, kesehatan masyarakat.

**CLUSTER ANALYSIS TO DETERMINE POTENTIAL DENGUE  
HEMORRHAGIC FEVER DISEASE AREAS  
IN WEST JAVA**

**Maulida Afifah**

**ABSTRACT**

*Dengue fever (DHF) is one of the serious public health problems in Indonesia, especially in West Java, the region with the highest DHF cases in 2022. With the background of increasing DHF cases associated with factors such as population density and high rainfall, this study applied the K-Means Clustering method to identify districts/cities in West Java based on DHF potential. Using data from 2020 to 2022 which includes population density, number of DHF cases, and flood events obtained from the West Java Government website. The analysis results show that the optimal K value is 4, which is supported by the elbow method graph that shows a slowing decrease in the Sum of Squared Errors (SSE) value after the number of clusters reaches 4. This study successfully grouped the region into four clusters: High Endemic is an area with high number of cases and population density, Moderate Endemic is an area with high number of cases and high number of floods, Potential is an area with rare dengue cases and low population density and flood frequency. Finally, Sporadic indicates the lowest number of dengue cases, population density, and the lowest number of floods. The visualization results are displayed in the form of an interactive map that makes it easier to understand the distribution of clusters in various regions in 2020-2022. These results are useful for directing health interventions and resources more effectively in West Java, as well as supporting data-driven decision making in dengue control.*

**Keywords:** *Dengue fever, K-Means clustering, data visualization, public health.*

## KATA PENGANTAR

Dengan penuh kerendahan hati, penulis mengucapkan puji syukur kepada Allah Swt atas segala nikmat dan hidayah-Nya yang telah memberikan kemudahan dan kelancaran sehingga penulis dapat menyelesaikan Skripsi ini. Penulisan Skripsi ini bertujuan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Dalam penyusunan Tugas Akhir Skripsi, Penulis menyadari bahwa dalam penyelesaian tugas akhir skripsi ini tentu tidak mungkin terpenuhi tanpa adanya dukungan, bimbingan, kritik dan saran dari berbagai pihak. mudah sehingga memerlukan dukungan dan arahan dari berbagai pihak. Maka penulis ucapkan rasa hormat dan terima kasih kepada:

1. Allah SWT yang telah memberikan kemudahan dalam segala proses, memberikan nikmat yang tiada henti, serta memberikan Rahmat dan karunia-Nya.
2. Kedua orang tua tercita. Bapa dan Ibu yang selalu mendoakan, memberi dukungan, memberikan kasih sayang serta nasihat yang menjadikan motivasi penulis dalam menyelesaikan skripsi ini.
3. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. Selaku Dosen Pembimbing 1 dan Pak Musthofa Galih Pradana, S.Kom., M.Kom. Selaku Dosen Pembimbing 2 yang telah bersedia meluangkan waktunya untuk membantu, membimbing, serta mengarahkan penulis dalam proses pengerjaan skripsi ini.
4. Bapak dan Ibu Dosen Fakultas Ilmu Komputer UPN Veteran Jakarta yang telah memberikan ilmu, bimbingan, serta arahan selama masa perkuliahan.
5. Adik-adik Penulis, Rayhan, Silmi, dan Sultan yang selalu memberikan dukungan dan selalu membantu penulis dalam hal apapun.
6. Keluarga Besar, Khususnya Mbah Kakung dan Mbah Biyung, yang senantiasa memberikan perhatian, doa, dan dukungan yang tiada hentinya.



7. Sahabat penulis, Fathma dan Amel, yang selalu setia memberikan dukungan, semangat, serta tempat berbagi keluh kesah.
8. Teman-teman seperjuangan Aulia, Jevani, Dinda, Alifia yang telah membantu, bekerja sama, dan memberi semangat untuk dapat menyelesaikan skripsi ini dengan baik dan lancar.
9. Teman-teman Komunitas Pelita Rumba yang telah memberikan bantuan, dorongan semangat, serta dukungan tak tergantikan selama proses penulisan skripsi ini.
10. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu per satu terimakasih atas do'a serta dukungan yang sangat berharga bagi penulis.

Penulis menyadari dalam penyusunan penelitian tugas akhir skripsi ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang akan dijadikan pembelajaran penulisan selanjutnya. Semoga skripsi ini dapat dipahami dan menambah pengetahuan bagi penulis sendiri maupun pembaca.

Jakarta, 28 April 2024

(Maulida Afifah)

## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iii
LEMBAR PERSETUJUAN.....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Ruang Lingkup.....	4
1.4. Tujuan.....	4
1.5. Manfaat.....	4
1.6. Luaran yang diharapkan .....	5
1.7. Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Demam Berdarah <i>Dengue</i> (DBD) .....	7
2.2 <i>Data mining</i> .....	9
2.3 Normalisasi Data .....	11
2.4 <i>Clustering</i> (Klasterisasi).....	12
2.5 Algoritma K-Means.....	13
2.6 <i>Metode Elbow</i> .....	14
2.7 <i>Python</i> .....	15
2.8 Streamlit Framework .....	16
2.9 Penelitian Yang Relevan .....	17

BAB III METODE PENELITIAN.....	22
3.1 Alur Penelitian.....	22
3.2 Identifikasi Masalah .....	23
3.3 Pengumpulan Data .....	23
3.4 Pemahaman Data.....	24
3.5 Pre-Processing Data .....	25
3.6 Penerapan Algoritma <i>K-Means</i> .....	25
3.7 Analisis dan Visualisasi Hasil <i>Clustering</i> .....	26
3.8 Alat Bantu Penelitian.....	27
3.9 Jadwal Penelitian .....	27
BAB IV HASIL DAN PEMBAHASAN .....	29
4.1 Pemahaman Data.....	29
4.2 <i>Preprocessing</i> Data .....	31
4.2.1. Pemilihan Atribut.....	31
4.2.2. Penggabungan Dataset .....	33
4.2.3. Normalisasi Data.....	35
4.3 <i>Clustering K-Means</i> .....	36
4.3.1 Menetapkan jumlah kluster (k) dengan Metode Elbow .....	36
4.3.2 Menentukan <i>centroid</i> .....	38
4.3.3 Mengukur jarak antara objek dengan <i>centroid</i> .....	38
4.3.4 Menghitung <i>centroid</i> baru.....	40
4.3.5 Mengulangi tahap dua sampai data pada tiap kluster terpusat .....	40
4.3.6 Hasil Perhitungan <i>Centroid</i> Terakhir .....	41
4.4 Analisis dan Visualisasi data .....	43
4.4.1 Analisis <i>Cluster</i> .....	44
4.4.2 Visualisasi Data.....	55
4.4.3 Tampilan Streamlit.....	64

BAB V PENUTUP.....	67
5.1    Kesimpulan.....	67
5.2    Saran.....	68
DAFTAR PUSTAKA.....	69
LAMPIRAN.....	72

## DAFTAR GAMBAR

Gambar 2.1 Contoh Gambar <i>Knowledge Discovery Database</i> (KDD) Sumber: Binus 2021 .....	10
Gambar 2. 2 Contoh Metode Elbow .....	15
Gambar 3. 1 Tahapan alur diagram.....	22
Gambar 3. 2 Jadwal Kegiatan Penelitian .....	28
Gambar 4. 1 Data Kasus DBD Berdasarkan Kabupaten/Kota.....	29
Gambar 4. 2 Data Kepadatan Penduduk Jawa Barat.....	30
Gambar 4. 3 Data Jumlah Kasus Banjir di Jawa Barat .....	30
Gambar 4. 4 Atribut data kasus DBD di Jawa Barat.....	32
Gambar 4. 5 atribut data kepadatan penduduk di Jawa Barat .....	32
Gambar 4. 6 atribut Data Jumlah Kejadian Bencana Banjir .....	33
Gambar 4. 7 Dataset hasil penggabungan atribut.....	34
Gambar 4. 8 Hasil Normalisasi .....	36
Gambar 4. 9 Grafik Metode Elbow.....	37
Gambar 4. 10 Hasil <i>Cluster centroid</i> awal.....	40
Gambar 4. 11 Hasil <i>Cluster</i> .....	42
Gambar 4. 12 Distribusi <i>cluster</i> .....	44
Gambar 4. 13 Box plot Jumlah Kasus DBD antar <i>cluster</i> .....	44
Gambar 4. 14 Box plot kepadatan penduduk antar klaster .....	45
Gambar 4. 15 Box plot jumlah banjir antar klaster.....	47
Gambar 4. 16 Filter Tahun .....	55
Gambar 4. 17 Filter nama_kabupaten_kota .....	56
Gambar 4. 18 kode filter .....	56
Gambar 4. 19 Filtrasi.....	57
Gambar 4. 20 Tabel filtrasi .....	57
Gambar 4. 21 plot garis berdasarkan kepadatan penduduk.....	58
Gambar 4. 22 Perubahan kasus DBD tahun 2020-2022 .....	60
Gambar 4. 23 Visualisasi Maps Wilayah Berpotensi DBD tahun 2020 .....	61
Gambar 4. 24 Visualisasi Maps Wilayah Berpotensi DBD tahun 2021 .....	62
Gambar 4. 25 Visualisasi Maps Wilayah Berpotensi DBD tahun 2022 .....	62
Gambar 4. 26 Tampilan Menu Analisis <i>Cluster</i> .....	64
Gambar 4. 27 Tampilan Menu Data Visualisasi Bagian 1 .....	65
Gambar 4. 28 Tampilan Menu Data Visualisasi Bagian 2.....	65

## DAFTAR TABEL

Tabel 2. 1 Penelitian yang Relevan.....	17
Tabel 4. 1 Atribut dataset potensi wilayah DBD .....	34
Tabel 4. 2 Skala Atribut-Atribut .....	35
Tabel 4. 3 normalisasi python .....	36
Tabel 4. 4 Hasil Selisih SSE .....	37
Tabel 4. 5 <i>Centroid</i> Awal.....	38
Tabel 4. 6 Hasil <i>Centroid</i> Akhir.....	41
Tabel 4. 7 Tabel Hasil Kabupaten/Kota terdekat dengan <i>centroid</i> akhir .....	41
Tabel 4. 8 <i>Clustering</i> K-Means Python.....	42
Tabel 4. 9 Presentase hasil <i>Clustering</i> .....	43
Tabel 4. 10 Wilayah <i>Cluster</i> 1 .....	48
Tabel 4. 11 Wilayah <i>Cluster</i> 2 .....	50
Tabel 4. 12 Wilayah <i>Cluster</i> 3 .....	52
Tabel 4. 13 Wilayah <i>Cluster</i> 4 .....	53