



**ANALISIS DAERAH POTENSI PERSEBARAN DEMAM  
BERDARAH *DENGUE* DI DKI JAKARTA DENGAN  
MENGUNAKAN METODE *CLUSTERING K-MEANS***

**SKRIPSI**

**HANI DZIKRA NURKHAIRIYYAH**

**1810512007**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**2022**



**ANALISIS DAERAH POTENSI PERSEBARAN DEMAM  
BERDARAH *DENGUE* DI DKI JAKARTA DENGAN  
MENGUNAKAN METODE *CLUSTERING K-MEANS***

**SKRIPSI**

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

**HANI DZIKRA NURKHAIRIYYAH**

**1810512007**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**2022**

## PERNYATAAN ORISINALITAS

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

Nama : Hani Dzikra Nurkhairiyah  
NIM : 1810512007  
Program Studi : S1 – Sistem Informasi  
Judul : Analisis Daerah Potensi Persebaran Demam Berdarah  
*Dengue* di DKI Jakarta dengan Menggunakan Metode  
*Clustering K-Means*  
Tanggal : 23 Juni 2022

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

Jakarta, 23 Juni 2022

Yang Menyatakan,



(Hani Dzikra Nurkhairiyah)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI  
UNTUK KEPENTINGAN AKADEMIS**

---

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Hani Dzikra Nurkhairiyah  
NIM : 1810512007  
Fakultas : Ilmu Komputer  
Program Studi : S1 – Sistem Informasi

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 DAERAH POTENSI PERSEBARAN DEMAM BERDARAH  
DENGUE DI DKI JAKARTA DENGAN MENGGUNAKAN METODE  
CLUSTERING K-MEANS**

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 : 23 Juni 2022

Yang Menyatakan,




(Hani Dzikra Nurkhairiyah)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Hani Dzikra Nurkhairiyah  
NIM : 1810512007  
Program Studi : S1 - Sistem Informasi  
Judul Skripsi : Analisis Daerah Potensi Persebaran Demam Berdarah  
*Dengue* di DKI Jakarta dengan Menggunakan Metode  
*Clustering K-Means*

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




**Dr. Ermatita, M.Kom.**  
Penguji I



**Ria Astriratma, S.Komp., M.Cs.**  
Penguji II




**Ati Zaidiah, S.Kom., M.TI.**  
Pembimbing I



**Helena Nurramdhani I. S.Pd., M.Kom.**  
Pembimbing II



**Dr. Ermatita, M.Kom.**  
Dekan Fakultas Ilmu Komputer



**Helena Nurramdhani I. S.Pd., M.Kom.**  
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 07 Juli 2022



# ANALISIS DAERAH POTENSI PERSEBARAN DEMAM BERDARAH *DENGUE* DI DKI JAKARTA DENGAN MENGGUNAKAN METODE *CLUSTERING K-MEANS*

Hani Dzikra Nurkhairiyah

## Abstrak

Demam Berdarah *Dengue* merupakan salah satu penyakit berjangkit di Indonesia. Penyakit demam berdarah *dengue* diakibatkan oleh gigitan nyamuk *Aedes aegypti* dan *Aedes Albopictus*. Penelitian ini bertujuan untuk melakukan *clustering* daerah potensi persebaran DBD di DKI Jakarta. *Clustering* dilakukan dengan menggunakan algoritma *K-Means*, dengan menentukan banyaknya *cluster* *K* yang akan dibagi. Data yang digunakan hasil dari perancangan data dengan menggabungkan beberapa atribut. Nilai *K* ditentukan melalui hasil evaluasi *Silhouette* dengan menghitung *Silhouette Score*. *Silhouette Score* yang dihasilkan sebesar 0.41642 pada  $K=2$ , sehingga diputuskan membagi data menjadi 2 *cluster*. Hasil *clustering* terdiri dari C0 yaitu daerah sporadis, meliputi 30 kecamatan, yaitu Cempaka Putih, Gambir, Johar Baru, Kemayoran, Menteng, Sawah Besar, Senen, Tanah Abang, Kelapa Gading, Koja, Pademangan, Tanjung Priok, Grogol Petamburan, Palmerah, Taman Sari, Tambora, Cilandak, Jagakarsa, Kebayoran Baru, Kebayoran Lama, Mampang Prapatan, Pancoran, Pesanggrahan, Setiabudi, Tebet, Matraman, Pasar Rebo, Pulo Gadung, Kepulauan Seribu Selatan, dan Kepulauan Seribu Utara. C1 daerah endemis terdiri dari 14 kecamatan, di antaranya Cilincing, Penjaringan, Cengkareng, Kalideres, Kebon Jeruk, Kembangan, Pasar Minggu, Cakung, Cipayung, Ciracas, Duren Sawit, Jatinegara, Kramat Jati, dan Makasar.

**Kata Kunci:** *clustering*, demam berdarah *dengue*, data mining, *k-means*

# **ANALISIS DAERAH POTENSI PERSEBARAN DEMAM BERDARAH *DENGUE* DI DKI JAKARTA DENGAN MENGGUNAKAN METODE *CLUSTERING K-MEANS***

**Hani Dzikra Nurkhairiyyah**

## ***Abstract***

*Dengue Hemorrhagic Fever is one of the infectious diseases in Indonesia. Dengue fever is caused by the bite of *Aedes aegypti* and *Aedes albopictus* mosquitoes. This study aims to cluster the potential areas for the distribution of DHF in DKI Jakarta. Clustering is done using the K-Means algorithm, by determining the number of k clusters to be divided. The data used is the result of data design by combining several attributes. The value of K is determined through the results of the Silhouette evaluation by calculating the Silhouette Score. The resulting Silhouette Score is 0.41642 at K=2, so it was decided to divide the data into 2 clusters. The clustering results consist of C0 is a sporadic area, include 30 sub-districts, namely Cempaka Putih, Gambir, Johar Baru, Kemayoran, Menteng, Sawah Besar, Senen, Tanah Abang, Kelapa Gading, Koja, Pademangan, Tanjung Priok, Grogol Petamburan, Palmerah, Taman Sari, Tambora, Cilandak, Jagakarsa, Kebayoran Baru, Kebayoran Lama, Mampang Prapatan, Pancoran, Pesanggrahan, Setiabudi, Tebet, Matraman, Pasar Rebo, Pulo Gadung, Kepulauan Seribu Selatan, dan Kepulauan Seribu Utara. C1 endemic area, consists of 14 sub-districts, which is Cilincing, Penjaringan, Cengkareng, Kalideres, Kebon Jeruk, Kembangan, Pasar Minggu, Cakung, Cipayung, Ciracas, Duren Sawit, Jatinegara, Kramat Jati, and Makasar.*

***Keywords:*** *clustering, dengue hemorrhagic fever, data mining, k-means*

## KATA PENGANTAR

Dengan memanjatkan puji dan syukur atas kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan judul “Analisis Daerah Potensi Persebaran Demam Berdarah *Dengue* di DKI Jakarta dengan Menggunakan Metode *Clustering K-means*”, guna memenuhi salah satu syarat untuk memperoleh gelar Sarjana Komputer pada program studi S1 Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Penulis menyadari bahwa tugas akhir ini dapat terselesaikan karena adanya bantuan dari berbagai pihak. Maka penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dr. Ermatita, M.Kom., selaku Dekan Fakultas Ilmu Komputer.
2. Ibu Helena Nurramdhani Irmada, S.Pd., M.Kom., selaku ketua program studi S1 Sistem Informasi serta dosen pembimbing 2 atas segala saran, arahan, dan bimbingan yang telah diberikan kepada penulis dalam menyelesaikan tugas akhir ini.
3. Ibu Ati Zaidiah, S.Kom., MTI., selaku dosen pembimbing 1 yang telah membimbing, memberikan arahan, dan saran kepada penulis sehingga tugas akhir ini dapat terselesaikan dengan baik.
4. Bapak Bambang Triwahyono, S.Kom., selaku dosen pembimbing akademik.
5. Ibu dan Bapak dosen Fakultas Ilmu Komputer yang telah memberikan ilmu pengetahuan yang bermanfaat.
6. Kedua orang tua yang selalu memberikan segenap kasih sayang, doa, dukungan, serta kesabarannya kepada penulis.
7. Teman-teman yang saling memberikan pengetahuan dan pemahaman. Serta teman terdekat penulis, Anissa Sari Dewi yang selalu memberikan dukungan.

Jakarta, 30 Mei 2022

Penulis



## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	ii
PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN.....	xiv
DAFTAR SIMBOL.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	4
1.5. Ruang Lingkup Penelitian .....	4
1.6. Luaran yang Diharapkan .....	5
1.7. Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penyakit Menular .....	7
2.2 Demam Berdarah <i>Dengue</i> (DBD) .....	8
2.3 <i>Data mining</i> .....	9
2.4 Metode Klastering ( <i>Clustering</i> ).....	11
2.4.1 Pengelompokan Hirarkis .....	11
2.4.2 Pengelompokan Partisi.....	12

2.5	Algoritma <i>K-means</i> .....	12
2.6	<i>Python</i> .....	14
2.7	Penelitian yang Relevan .....	15
BAB III METODOLOGI PENELITIAN.....		20
3.1	Alur Penelitian.....	20
3.2	Identifikasi Masalah .....	21
3.3	Metode Pengumpulan Data .....	21
3.4	Pemahaman Data .....	22
3.5	Perancangan <i>Dataset</i> .....	23
3.6	Pengolahan Data Menggunakan Algoritma <i>K-means</i> .....	24
3.7	Alat Bantu Penelitian.....	24
3.8	Jadwal Penelitian .....	25
BAB IV HASIL DAN PEMBAHASAN .....		27
4.1	Pemahaman Data.....	27
4.2	Perancangan <i>Dataset</i> .....	31
4.3	Perhitungan Manual Algoritma <i>K-Means</i> .....	36
4.4	Pengolahan Data Menggunakan Algoritma <i>K-Means</i> .....	51
4.5	Analisis Hasil <i>Clustering</i> & Visualisasi Data .....	54
BAB V PENUTUP.....		65
5.1	Kesimpulan.....	65
5.2	Saran .....	65
DAFTAR PUSTAKA .....		66
RIWAYAT HIDUP.....		70
LAMPIRAN.....		71

## DAFTAR TABEL

Tabel 1. Proses Algoritma <i>K-means</i> .....	13
Tabel 2. Penelitian yang Relevan.....	15
Tabel 3. Jadwal Penelitian.....	25
Tabel 4. Atribut Data Jumlah Penderita DBD .....	28
Tabel 5. Atribut Data Jumlah Kependudukan.....	29
Tabel 6. Atribut Data Kejadian Banjir .....	30
Tabel 7. <i>Dataset</i> Lengkap Hasil Normalisasi .....	37
Tabel 8. Nilai <i>Centroid</i> Awal.....	38
Tabel 9. Hasil <i>Cluster</i> Iterasi Ke-1 .....	39
Tabel 10. Hasil <i>Centroid</i> Baru .....	41
Tabel 11. Hasil <i>Clustering</i> Iterasi ke-2 .....	41
Tabel 12. Hasil <i>Centroid</i> Baru Iterasi ke-2 .....	42
Tabel 13. Hasil <i>Clustering</i> Iterasi ke-3 .....	43
Tabel 14. Hasil <i>Centroid</i> Baru Iterasi ke-3 .....	44
Tabel 15. Hasil <i>Clustering</i> Iterasi ke-4 .....	44
Tabel 16. Hasil <i>Centroid</i> Baru Iterasi ke-4 .....	45
Tabel 17. Hasil <i>Clustering</i> Iterasi ke-5 .....	46
Tabel 18. Hasil <i>Centroid</i> Baru Iterasi ke-5 .....	47
Tabel 19. Hasil <i>Clustering</i> Iterasi ke-6 .....	47
Tabel 20. Hasil <i>Centroid</i> Baru Iterasi ke-6 .....	48
Tabel 21. Hasil <i>Clustering</i> Iterasi ke-7 .....	49
Tabel 22. Hasil <i>Centroid</i> Baru Iterasi ke-7 .....	50
Tabel 23. Hasil <i>Final Clustering</i> .....	50

## DAFTAR GAMBAR

Gambar 1. Kepadatan Penduduk DKI Jakarta Tahun 2020 .....	2
Gambar 2. Penyakit-penyakit yang Menyebabkan Wabah .....	8
Gambar 3. <i>Hierarchical Clustering</i> .....	12
Gambar 4. Proses <i>Clustering</i> Obyek dengan Teknik <i>K-means</i> .....	12
Gambar 5. Alur Penelitian.....	20
Gambar 6. Data Penderita DBD.....	27
Gambar 7. Data Kepadatan Penduduk DKI Jakarta.....	28
Gambar 8. Data Kejadian Bencana Banjir di DKI Jakarta Bulan Januari .....	29
Gambar 9. Atribut Terpilih pada Data Penderita DBD.....	32
Gambar 10. Atribut Terpilih pada Data Kepadatan Penduduk DKI Jakarta.....	32
Gambar 11. Atribut Terpilih pada Data Kejadian Banjir di DKI Jakarta .....	33
Gambar 12. <i>Dataset</i> Penyebaran DBD (diolah).....	34
Gambar 13. Data Teratas <i>Dataset</i> .....	35
Gambar 14. Hasil Seleksi Data .....	35
Gambar 15. <i>Source Code</i> Normalisasi <i>Min-Max</i> .....	36
Gambar 16. <i>Dataset</i> Hasil Normalisasi.....	36
Gambar 17. <i>Source Code</i> Penerapan <i>K-Means</i> .....	51
Gambar 18. Hasil <i>Clustering</i> dengan Nilai $K=3$ .....	52
Gambar 19. <i>Silhouette Score</i> .....	52
Gambar 20. <i>Source Code</i> Penerapan <i>K-Means</i> $K=2$ .....	53
Gambar 21. <i>DataFrame</i> Hasil <i>Clustering</i> .....	53
Gambar 22. Histogram Anggota <i>Cluster</i> .....	53
Gambar 23. Tampilan <i>DataFrame</i> Hasil <i>Cluster</i> .....	54
Gambar 24. <i>Boxplot Cluster</i> dan Penderita DBD .....	54
Gambar 25. <i>Boxplot Cluster</i> dan Luas Wilayah .....	55
Gambar 26. <i>Boxplot Cluster</i> dan Kepadatan Penduduk.....	55
Gambar 27. <i>Boxplot Cluster</i> dan RW Terdampak Banjir .....	56
Gambar 28. <i>Boxplot Cluster</i> dan RT Terdampak Banjir.....	56
Gambar 29. Visualisasi Hasil <i>Cluster</i> .....	57



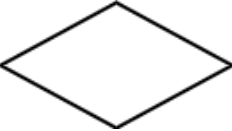

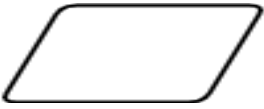

Gambar 30. Visualisasi <i>Seaborn</i> Baris ke-5 .....	58
Gambar 31. Visualisasi <i>Seaborn</i> Baris ke-4 .....	59
Gambar 32. Visualisasi <i>Seaborn</i> Baris ke-3 .....	60
Gambar 33. Visualisasi <i>Seaborn</i> Baris ke-2 .....	60
Gambar 34. Visualisasi <i>Seaborn</i> Baris ke-1 .....	61
Gambar 35. Visualisasi <i>Maps Cluster</i> Daerah Potensi Persebaran DBD di DKI Jakarta .....	63

## DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i> Penderita DBD di Provinsi DKI Jakarta awal.....	72
Lampiran 2 <i>Dataset</i> Kepadatan Penduduk di DKI Jakarta awal.....	81
Lampiran 3 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Januari.....	85
Lampiran 4 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Februari.....	87
Lampiran 5 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Maret.....	89
Lampiran 6 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan April.....	90
Lampiran 7 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Mei.....	90
Lampiran 8 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Juni.....	91
Lampiran 9 <i>Dataset</i> Kejadian Banjir di DKI Jakarta Bulan Juli .....	91
Lampiran 10 <i>Dataset</i> Kejadian Banjir Bulan Agustus.....	91
Lampiran 11 <i>Dataset</i> Kejadian Banjir Bulan September.....	91
Lampiran 12 <i>Dataset</i> Kejadian Banjir Bulan Oktober.....	92
Lampiran 13 <i>Dataset</i> Kejadian Banjir Bulan November.....	96
Lampiran 14 <i>Dataset</i> Kejadian Banjir Bulan Desember.....	97
Lampiran 15 <i>Dataset</i> yang Sudah diolah.....	98
Lampiran 16 <i>Source Code Import Library</i> .....	99
Lampiran 17 <i>Source Code</i> Menampilkan <i>Dataset</i> .....	99
Lampiran 18 <i>Source Code</i> Seleksi Data.....	99
Lampiran 19 <i>Source Code</i> Membuat <i>Data Frame</i> Baru Hasil <i>Min-Max</i> .....	99
Lampiran 20 <i>Source Code</i> Menampilkan Hasil <i>CLustering</i> .....	99
Lampiran 21 <i>Source Code</i> Mencari <i>Silhouette Score</i> .....	99
Lampiran 22 <i>Source Code</i> Menambahkan <i>Feature Cluster</i> .....	100
Lampiran 23 <i>Source Code</i> Menampilkan Histogram.....	100
Lampiran 24 <i>Source Code</i> Menambahkan Kembali <i>Feature</i> Kecamatan .....	100
Lampiran 25 <i>Source Code</i> Menampilkan <i>Boxplot</i> Antara <i>Cluster</i> dan Penderita DBD.....	100
Lampiran 26 <i>Source Code</i> Menampilkan <i>Boxplot</i> Antara <i>Cluster</i> dan Luas Wilayah.....	100

Lampiran 27 <i>Source Code</i> Menampilkan <i>Boxplot</i> Antara <i>Cluster</i> dan Kepadatan Penduduk .....	100
Lampiran 28 <i>Source Code</i> Menampilkan <i>Boxplot</i> Antara <i>Cluster</i> dan RW Terdampak Banjir.....	100
Lampiran 29 <i>Source Code</i> Menampilkan <i>Boxplot</i> Antara <i>Cluster</i> dan RT Terdampak Banjir .....	101
Lampiran 30 <i>Source Code</i> Menampilkan Visualisasi <i>Seaborn</i> .....	101
Lampiran 31 Hasil Turnitin.....	102

## DAFTAR SIMBOL

No.	Simbol	Nama Simbol	Keterangan
1.		<i>Flow</i>	Sebagai arah alur dalam prosedur yang mengarahkan ke instruksi selanjutnya.
2.		<i>Process</i>	Proses yang dilakukan.
3.		<i>Decision</i>	Untuk menunjukkan sebuah situasi yang berbeda sesuai dengan keputusan yang digambarkan.
4.		<i>Terminator</i>	Awal dan akhir dari proses.
5.		Data	Menggambarkan <i>input</i> dan <i>output</i> .
6.		<i>Document</i>	Suatu masukan dari dokumen.