

**"Pengembangan Fungsi R untuk Analisis Sebaran Spasial-Temporal dan Clustering
Hotspot Karhutla di Provinsi Sumatera Selatan (Januari - Juli 2025)"**

TUGAS BESAR MATA KULIAH KOMPUTASI STATISTIKA



Kelompok 3
Disusun oleh :

Luthfia Laila Ramadhani	123450004
Feby Angelina	123450039
Ginda Fajar Riadi Marppaung	123450103
Muhammad Dzikra	123450124

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2025**

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan fungsi terpadu dalam bahasa pemrograman R guna mengotomatisasi analisis spasial-temporal dan *clustering* titik panas (*hotspot*) kebakaran hutan dan lahan di Provinsi Sumatera Selatan periode Januari hingga Juli 2025. Menggunakan data sekunder dari sensor VIIRS satelit NOAA-20 yang memuat 2.793 observasi tervalidasi, metodologi penelitian ini mengintegrasikan tahapan *data wrangling*, transformasi spasial, analisis tren temporal, serta penerapan algoritma *K-Means Clustering* untuk zonasi wilayah rawan. Stabilitas hasil klaster diuji menggunakan metode validasi *resampling bootstrap* dengan 100 iterasi. Hasil utama menunjukkan adanya pola peningkatan eksponensial *hotspot* yang memuncak pada bulan Juli dengan 1.404 titik kejadian, serta pola deteksi harian yang terkonsentrasi pada pukul 06:00 UTC. Secara spasial, terbentuk tiga klaster karakteristik, di mana Klaster 1 (Pesisir Timur) mencatat intensitas kebakaran tertinggi dengan rata-rata *Fire Radiative Power* (FRP) 8,73 MW, sedangkan Klaster 3 (Tengah-Selatan) memiliki frekuensi kejadian tertinggi. Validasi model menghasilkan nilai *Coefficient of Variation* (CV) sebesar 1,80%, mengindikasikan struktur klaster yang sangat stabil. Kesimpulannya, pengembangan fungsi komputasi ini terbukti efektif menyediakan alur kerja otomatis untuk pemetaan risiko karhutla yang akurat dan dapat diandalkan bagi pemangku kebijakan.

Kata kunci : Spasial-Temporal, Hotspot, Karhutla, K-Means Clustering, Pemrograman R.

ABSTRACT

This study aims to develop an integrated function within the R programming language to automate the spatial-temporal analysis and clustering of forest and land fire (karhutla) hotspots in South Sumatra Province from January to July 2025. Utilizing secondary data from the NOAA-20 satellite's VIIRS sensor containing 2,793 validated observations , the research methodology integrates stages of data wrangling, spatial transformation, temporal trend analysis, and the implementation of the K-Means Clustering algorithm for vulnerable area zoning. Cluster stability was evaluated using the bootstrap resampling validation method with 100 iterations. Key findings reveal an exponential increase pattern in hotspots, peaking in July with 1,404 incidents, as well as a diurnal detection pattern concentrated at 06:00 UTC. Spatially, three distinct clusters were identified: Cluster 1 (East Coast) recorded the highest fire intensity with an average Fire Radiative Power (FRP) of 8.73 MW, while Cluster 3 (Central-South) exhibited the highest occurrence frequency. Model validation yielded a Coefficient of Variation (CV) of 1.80%, indicating a highly stable cluster structure. In conclusion, the development of this computational function demonstrates effectiveness in providing an automated workflow for accurate and reliable forest fire risk mapping to support policymakers.

Keywords : Spatial-Temporal, Hotspot, Forest and Land Fires, K-Means Clustering, R Programming.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hutan hujan tropis Indonesia merupakan ekosistem vital yang menjaga keseimbangan lingkungan global melalui fungsi hidrologis, penyediaan oksigen, dan penyerapan karbon. Namun, fungsi ekologis ini terancam oleh tantangan besar berupa kebakaran hutan dan lahan (karhutla) yang terjadi setiap tahun. Fenomena ini membawa dampak merugikan, mulai dari kerusakan siklus air, penurunan kualitas tanah, hingga hilangnya vegetasi penutup. Statistik menunjukkan bahwa hampir 99% kejadian karhutla disebabkan oleh aktivitas manusia, baik disengaja maupun tidak[1]. Kondisi ini semakin diperparah oleh cuaca ekstrem akibat fenomena El Nino pada musim kemarau serta kondisi lahan yang telah terdegradasi. Di Indonesia, Sumatera Selatan menjadi salah satu wilayah yang menghadapi ancaman paling serius dari permasalahan ini. Dampak karhutla di wilayah tersebut tidak hanya merusak ekosistem, tetapi juga meluas pada gangguan kesehatan masyarakat, kerugian ekonomi, dan penurunan kualitas udara regional secara signifikan.

Data hotspot dari satelit VIIRS (*Visible Infrared Imaging Radiometer Suite*) yang dioperasikan oleh NOAA-20 dapat digunakan untuk memantau dan menganalisis pola sebaran titik api[8]. Namun, data mentah ini memerlukan proses analisis yang kuat untuk ditindaklanjuti menjadi informasi yang bermanfaat. Pemahaman terhadap pola spasial dan temporal hotspot sangat penting untuk mengidentifikasi wilayah rawan kebakaran dan merancang strategi mitigasi yang efektif. Analisis clustering dapat membantu mengelompokkan hotspot berdasarkan karakteristik geografisnya, sehingga dapat diketahui zona-zona konsentrasi kebakaran. Validasi bootstrap diperlukan untuk memastikan stabilitas hasil clustering dan keandalan pola yang ditemukan

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan fungsi khusus dalam bahasa R yang mampu melakukan analisis sebaran spasial-temporal dan clustering hotspot karhutla di Provinsi Sumatera Selatan untuk periode Januari hingga Juli 2025. Fungsi yang dikembangkan akan mengintegrasikan langkah-langkah mulai dari impor data, pembersihan, analisis statistik deskriptif temporal, visualisasi sebaran spasial, hingga penerapan algoritma *clustering* seperti K-Means untuk mengidentifikasi area dengan konsentrasi hotspot yang tinggi. Dengan adanya fungsi yang terotomasi ini, diharapkan dapat menjadi alat yang efisien bagi pemerintah daerah, BNPB, dan peneliti untuk memetakan dinamika karhutla, mengidentifikasi zona rawan berulang, dan pada akhirnya mendukung pengambilan keputusan yang lebih cepat dan berbasis data dalam upaya mitigasi bencana karhutla.

1.2 Rumusan Masalah

1. Bagaimana merancang sebuah fungsi R yang mampu mengotomatisasi proses data wrangling dan analisis spasial-temporal (distribusi bulanan) dari data hotspot VIIRS mentah?
2. Bagaimana mengimplementasikan algoritma K-Means Clustering ke dalam fungsi R untuk mengidentifikasi zona konsentrasi hotspot secara otomatis dan efisien?
3. Bagaimana mengintegrasikan metode *resampling bootstrap* di dalam fungsi R untuk mengukur stabilitas dan validitas hasil *clustering*?
4. Bagaimana mengembangkan *output* dari fungsi tersebut agar dapat menghasilkan visualisasi data (peta dan grafik) dan ringkasan statistik yang informatif bagi pengguna?

1.3 Tujuan

1. Membangun sebuah fungsi R yang dapat memproses data hotspot VIIRS untuk menganalisis distribusi spasial dan temporal secara otomatis di wilayah Sumatera Selatan.
2. Mengimplementasikan algoritma K-Means Clustering sebagai fungsi R untuk melakukan pengelompokan hotspot berdasarkan koordinat geografis.
3. Mengintegrasikan metode validasi *resampling bootstrap* ke dalam fungsi untuk menilai stabilitas klaster yang ditemukan.
4. Menghasilkan sebuah fungsi R yang *output*-nya mencakup ringkasan statistik dan visualisasi data yang relevan.

BAB II

TINJAUAN PUSTAKA

2.1 Kebakaran Hutan dan Lahan di Indonesia

Indonesia memiliki lahan gambut yang cukup luas, menjadi keempat terbesar setelah Kanada, Rusia, dan Amerika Serikat. Luas total lahan gambut di Indonesia mencapai 13,43 juta hektar, terdapat di sepanjang pulau Sumatra (43,18%), Kalimantan (32,06%), dan Papua (24,76%). Namun, data sampai tahun 2022 menunjukkan hanya sekitar 16% dari total lahan gambut yang masih dalam kondisi tidak rusak, sekitar 4 juta hektar. Sedangkan jumlah lahan gambut yang terdegradasi meningkat tajam. Terdapat 206.935 hektar lahan gambut dalam kondisi rusak sangat berat karena sudah berubah fungsi menjadi lahan pertanian, perkebunan, dan HTI, serta menjadi prioritas restorasi.

Provinsi Sumatera Selatan memiliki luas lahan gambut sebesar 2,09 juta hektar, merupakan yang terbesar kedua di Pulau Sumatera setelah Riau. Provinsi ini pernah mengalami kebakaran hutan terparah pada tahun 2015 dengan luas 646.298 hektar dan pada tahun 2019 terjadi kebakaran sebanyak 336.798 hektar, sehingga menjadi salah satu provinsi yang rentan terhadap bencana karhutla yang ditetapkan oleh pemerintah. Dalam periode 2018-2021, total lahan yang terdegradasi akibat karhutla mencapai 462.627 hektar. Kabupaten Ogan Komering Ilir (OKI) memiliki luas lahan gambut terbesar di Sumatera Selatan, yaitu 1,03 juta hektar atau 49,28% dari total lahan gambut provinsi tersebut. Angka kejadiannya karhutla di OKI merupakan yang tertinggi di Sumatera Selatan pada tahun 2017-2018 dan 2021. Pada tahun 2015 sekitar 53,4% lahan gambut di OKI terbakar, dan pada tahun 2019 kembali terjadi 41% lahan terbakar. Dalam periode 2016-2018, karhutla tetap terjadi meski tidak dipicu oleh fenomena El Nino. Hal ini menunjukkan bahwa aktivitas manusia memiliki peran besar dalam terjadinya kebakaran di lahan gambut. Hingga saat ini, tercatat 116 hektar lahan di OKI yang terbakar sejak Januari hingga Agustus 2022. Angka ini sementara menjadi yang tertinggi di Sumatera Selatan. Hal ini menunjukkan bahwa upaya mitigasi yang dilakukan pemerintah serta masyarakat seharusnya dievaluasi karena kejadian serupa terus terjadi bahkan cenderung meningkat setiap tahunnya [2].

2.2 Sebaran Titik Panas (Hotspot)

Hotspot atau titik panas adalah istilah untuk sebuah piksel yang memiliki suhu lebih tinggi dari batas tertentu pada hasil interpretasi gambar satelit, yang bisa menjadi tanda adanya kebakaran hutan atau lahan. Untuk menganalisis data titik panas, digunakan data dengan tingkat kepercayaan minimal 50%. Titik panas sering ditemukan pada bulan yang memiliki iklim kering. Pada masa bulan kering ini, masyarakat serta perusahaan kehutanan dan per-

kebunan biasanya melakukan aktivitas membersihkan lahan dengan cara membakar. Setelah itu, distribusi titik panas mengalami penurunan dari bulan September hingga Desember, karena kondisi iklim mulai memasuki musim hujan yang tinggi [3].

2.3 K-Means Clustering untuk Analisis Spasial

Algoritma K-Means *clustering* merupakan algoritma partisi yang paling banyak digunakan untuk analisis data multivariat, termasuk dalam konteks analisis data spasial. Algoritma ini bekerja secara iteratif dengan menugaskan titik data ke klaster terdekat dan memperbarui *centroid* (pusat klaster) hingga klaster mencapai kestabilan atau *convergence*[4]. Dalam konteks analisis *hotspot* geografis, K-Means terbukti efektif untuk mengidentifikasi konsentrasi spasial dan mengelompokkan lokasi berdasarkan kedekatan geografis [5].

Kelebihan K-Means meliputi efisiensi komputasi bahkan untuk *dataset* besar, serta kesederhanaan dalam implementasi dan interpretasi [6]. Algoritma ini memiliki keterbatasan, yaitu sensitivitas terhadap penempatan *centroid* awal yang dapat menyebabkan solusi optimal lokal. Kendala ini sering diatasi dengan menjalankan *multiple random starts*. Keterbatasan lain adalah keharusan untuk menentukan jumlah klaster (*k*) di awal, yang sering memerlukan *domain knowledge* atau teknik validasi klaster tambahan [4]. Dalam aplikasi untuk *hotspot clustering*, koordinat geografis digunakan sebagai ruang fitur. Euclidean *distance* dalam ruang koordinat ini merepresentasikan kedekatan geografis yang wajar untuk analisis skala regional [5]. Secara matematis, *Between-cluster sum of squares* dan *within-cluster sum of squares* menjadi matrik utama evaluasi, di mana rasio variansi antar-klaster terhadap total variasi mengindikasikan proporsi variasi data yang berhasil dijelaskan oleh proses klasterisasi tersebut [4].

2.4 Bootstrap Validation untuk Cluster Stability

Bootstrap adalah teknik pengambilan sampel ulang yang dikembangkan oleh Bradley Efron untuk mengestimasi distribusi sampel dan kuantifikasi ketidakpastian. Dalam konteks klustering, validasi *bootstrap* digunakan untuk menilai stabilitas solusi klaster terhadap gangguan atau perubahan kecil dalam data. Metode ini melibatkan pengambilan sampel ulang berulang dengan pengembalian dari dataset asli, melakukan klasterisasi pada setiap sampel *bootstrap*, dan memeriksa variasi hasil klasterisasi antar sampel [7].

Matrik stabilitas yang umum digunakan termasuk Coefficient of Variation (CV) dari jumlah kuadrat dalam klaster (Rizzo, 2019). Nilai CV rendah mengindikasikan stabilitas tinggi, yang berarti struktur klaster kuat (*robust*) dan tidak terlalu bergantung pada titik data tertentu. Menurut literatur validasi klasterisasi, $CV < 5\%$ umumnya dianggap sangat stabil, $CV 5\text{--}10\%$ sebagai stabil, dan $CV > 20\%$ mengindikasikan klasterisasi tidak stabil yang sensitif terhadap variasi sampel.

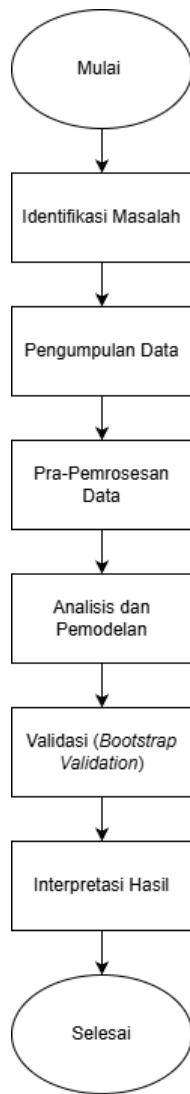
Validasi *bootstrap* sangat penting untuk analisis klaster karena beberapa alasan. Pertama, *clustering* adalah pembelajaran tanpa label (*unsupervised learning*) tanpa kebenaran acuan,

sehingga stabilitas menjadi kriteria penting untuk kualitas solusi. Kedua, data spasial sering kali memiliki autokorelasi spasial yang melekat dan heterogenitas yang dapat mempengaruhi hasil klasterisasi. Ketiga, pengambilan keputusan praktis memerlukan keyakinan bahwa klaster yang teridentifikasi benar-benar merepresentasikan pola spasial nyata dan bukan artefak statistik.

BAB III

METODE PENELITIAN

3.1 Diagram Alur Metode Penelitian



Gambar 1. Diagram Alir

3.2 Sumber Data

Dataset yang digunakan dalam penelitian ini bersumber dari NASA FIRMS (*Fire Information for Resource Management System*), sebuah platform penyedia data titik panas (*hotspot*) secara *near real-time* yang dioperasikan oleh NASA. Data *hotspot* diperoleh dari sensor VIIRS yang terpasang pada satelit NOAA-20, bagian dari *Joint Polar Satellite System* (JPSS) yang mengorbit bumi dalam lintasan *sun-synchronous* dengan kemampuan melintas di wilayah yang sama dua kali sehari. Periode observasi mencakup tujuh bulan, mulai dari 2 Januari hingga 31 Juli 2025,

yang menghasilkan total 3.072 entri deteksi *hotspot* awal. Data tersedia dalam dua format yang saling melengkapi yaitu CSV (*Comma-Separated Values*) yang memuat atribut tabular, dan *Shapefile* yang menyediakan representasi geografis dalam format vektor yang kompatibel dengan Sistem Informasi Geografis (SIG).

Setiap rekaman data dilengkapi dengan variabel komprehensif, mencakup koordinat geografis (*latitude* dan *longitude* dalam sistem WGS84/EPSC:4326), tingkat kecerahan termal (*brightness* dalam Kelvin), serta *Fire Radiative Power* (FRP) dalam satuan Megawatt yang merepresentasikan intensitas energi api. Informasi tersebut dilengkapi dengan *confidence level* (dikategorikan *low*, *nominal*, atau *high*) serta penanda waktu akuisisi UTC dan label *day/night*. Secara teknis, kualitas data VIIRS NOAA-20 sangat andal karena memiliki resolusi spasial 375 meter pada nadir, jauh lebih tinggi dibandingkan sensor MODIS generasi sebelumnya yang memiliki resolusi 1 kilometer [8]. Selain itu, resolusi temporal dengan dua kali lintas (*pass*) per hari dan latensi data kurang dari 3 jam memungkinkan deteksi frekuensi tinggi dan pemantauan *near real-time* yang krusial untuk analisis tanggap darurat.

3.3 Teknik Analisis

3.3.1 Data Preprocessing

Tahapan pra-pemrosesan data merupakan pondasi penting untuk memastikan kualitas dan konsistensi sebelum analisis dilakukan. Proses dimulai dengan pemuatan data (*data loading*) menggunakan paket *sf* untuk data spasial dan *dplyr* untuk manipulasi data. Paket *sf* menyediakan kerangka kerja modern untuk pengolahan data vektor spasial menggunakan standar *Simple Features* yang kompatibel dengan ekosistem *tidyverse*.

Penyaringan data (*data filtering*) diterapkan untuk memfokuskan analisis pada titik panas (*hotspot*) yang signifikan dengan ambang batas *Fire Radiative Power* (FRP) minimal 1 MW. Ambang batas ini dipilih berdasarkan literatur yang menunjukkan bahwa FRP di bawah 1 MW sering kali merepresentasikan kebakaran kecil atau deteksi palsu (*false positives*) dari sumber panas non-api seperti fasilitas industri atau pantulan sinar matahari (*sun glint*). Penyaringan ini mereduksi jumlah observasi dari 3.072 menjadi 2.793 titik panas yang lebih andal untuk dianalisis.

Rekayasa fitur (*feature engineering*) dilakukan secara ekstensif untuk memperkaya dataset dengan variabel temporal. Dari data mentah waktu akuisisi, diekstraksi komponen seperti bulan, tahun, jam, dan hari ke-n dalam tahun (*day of year*). Variabel penanda (*flag*) *is_night* diturunkan dari variabel asli *day_night* untuk klasifikasi biner antara deteksi siang dan malam hari.

Konversi spasial merupakan tahap krusial untuk mentransformasi data tabular menjadi objek spasial. Fungsi *st_as_sf()* digunakan untuk membuat objek *simple features* dengan sistem referensi koordinat WGS84 (EPSC:4326). Parameter koordinat menspesifikasikan kolom *longitude* dan *latitude* sebagai geometri, sementara kolom asli tetap dipertahankan untuk

keperluan analisis non-spasial. Objek spasial yang dihasilkan kompatibel dengan berbagai operasi spasial seperti penggabungan spasial (*spatial joins*), analisis *buffer*, dan perhitungan jarak.

3.3.2 Klastering K-Means (K-Means Clustering)

K-Means *clustering* dipilih sebagai metode utama untuk pengelompokan spasial karena efisiensi komputasi dan interpretabilitasnya yang baik pada data berbasis koordinat. Algoritma ini bekerja dengan mempartisi dataset menjadi k klaster yang tidak tumpang tindih berdasarkan metrik kesamaan (*similarity metric*).

Proses klasterisasi dimulai dengan ekstraksi koordinat geografis (*longitude* dan *latitude*) yang menjadi ruang fitur. Diasumsikan bahwa *Euclidean distance* dalam ruang koordinat merepresentasikan kedekatan geografis yang wajar untuk analisis skala regional seperti Sumatera Selatan, di mana distorsi akibat kelengkungan bumi relatif minimal.

Algoritma bekerja secara iteratif melalui dua langkah utama: *assignment* (penugasan titik ke *centroid* terdekat) dan *update* (perhitungan ulang posisi *centroid*). Iterasi berhenti ketika konvergensi tercapai. Parameter jumlah klaster ditetapkan $k=3$ berdasarkan pertimbangan geografis Sumatera Selatan yang terbagi menjadi tiga zona karakteristik: pesisir timur (gambut), zona barat (pegunungan Bukit Barisan), dan zona tengah (dataran rendah). Parameter *nstart* ditetapkan sebesar 25 untuk melakukan inisialisasi acak berulang guna menghindari solusi optimum lokal, dengan *random seed* 123 untuk menjamin reproduksibilitas.

Evaluasi kualitas klaster menggunakan metrik *Between-cluster Sum of Squares* (variabilitas antar-klaster) dan *Within-cluster Sum of Squares* (kekompakan dalam klaster). Rasio yang tinggi antara keduanya mengindikasikan pemisahan klaster yang baik.

3.3.3 Validasi Bootstrap (Bootstrap Validation)

Validasi bootstrap merupakan teknik resampling yang digunakan untuk menilai stabilitas dan keandalan solusi klaster. Metode ini sangat bernilai dalam analisis klaster karena ketiadaan label kebenaran dasar (*ground truth*) pada metode unsupervised learning.

Prosedur dimulai dengan pembuatan 100 sampel *bootstrap* dari dataset asli. Setiap sampel dibuat melalui pengambilan acak dengan pengembalian, sehingga ukuran sampel tetap sama (2.793 observasi) namun komposisinya bervariasi. Pada setiap sampel, algoritma K-Means dijalankan ulang (dengan parameter $k=3$, *nstart*=10). Nilai *Within-Cluster Sum of Squares* (WSS) diekstraksi dari setiap iterasi sebagai ukuran kekompakan klaster.

Distribusi dari 100 nilai WSS kemudian dianalisis. Stabilitas diukur menggunakan *Coefficient of Variation* (CV), yaitu rasio standar deviasi terhadap rata-rata. Interpretasi stabilitas mengacu pada panduan literatur: $CV < 5\%$ dianggap "Sangat Stabil", $CV 5\text{--}10\%$ "Stabil", dan $CV > 20\%$

"Tidak Stabil". Selain itu, distribusi *centroid* antar sampel *bootstrap* digunakan untuk mengkonstruksi interval kepercayaan guna memvisualisasikan ketidakpastian lokasi klaster.

3.4 Perangkat Lunak dan Lingkungan Komputasi

Analisis dilakukan menggunakan lingkungan pemrograman statistik R versi 4.3.3 . R dipilih karena merupakan platform sumber terbuka (*open-source*) dengan ekosistem pustaka yang luas untuk analisis spasial dan statistik. Kode didokumentasikan dalam format R Markdown untuk mengintegrasikan kode, hasil, dan narasi guna menjamin reproduksibilitas.

Pustaka utama yang digunakan meliputi:

1. *sf (Simple Features)*: Untuk penanganan data vektor spasial modern yang kompatibel dengan standar ISO 19125, menangani operasi geometri, I/O format, dan transformasi proyeksi (PROJ).
2. *dplyr*: Untuk manipulasi data (*data manipulation*) yang efisien menggunakan operator *pipe*, memfasilitasi transformasi dan agregasi data.
3. *ggplot2 & ggspatial*: Untuk visualisasi data statis berkualitas publikasi dan penambahan elemen spasial peta.
4. *lubridate*: Untuk mempermudah manipulasi data waktu (*date-time*), termasuk penanganan zona waktu.
5. *leaflet*: Untuk pembuatan peta web interaktif yang memungkinkan eksplorasi pola spasial secara dinamis.
6. *cluster*: Menyediakan implementasi algoritma klasterisasi dan alat diagnostik validasi.

Komputasi dijalankan pada lingkungan yang kompatibel dengan sistem operasi Windows/MacOS/Linux, dengan kebutuhan perangkat keras minimum RAM 8GB untuk menangani beban komputasi iteratif selama proses validasi *bootstrap*.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Dekripsi Dataset

Dataset yang digunakan dalam pengerjaan tugas besar kali ini yaitu data satelit yang berisi deteksi titik panas (*hotspot*) atau kebakaran di Sumatera Selatan berdasarkan koordinatnya sepanjang Januari hingga Mei 2025. File ini berisi kumpulan data observasi satelit VIIRS (*Visible Infrared Imaging Radiometer Suite*) yang diambil dari satelit N20 (NOAA-20). Berikut dataset yang kami gunakan:  fire_archive_J1V-C2_687296

Tabel 1. Statistik Deskriptif Variabel Utama

Variable	Mean	Min	Max	Std Dev
Latitude	-2.721	-4.499	-1.535	0.839
Longitude	104.236	102.001	106.49	1.445
Brightness (K)	334.15	295.30	367.00	12.48
FRP (MW)	7.46	1.00	73.70	6.92

Parameter utama yang akan dianalisis diantaranya:

1. Latitude dan Longitude merupakan koordinat geografis titik panas. Kolom ini merupakan data terpenting untuk pemetaan.
2. acq_date & acq_time merupakan waktu deteksi panas.
3. *Fire Radiative Power* (frp) merupakan daya radiasi api dalam satuan Megawatt. FRP merupakan indikator terbaik untuk mengukur intensitas atau kekuatan kebakaran. Nilai FRP yang tinggi berarti api lebih besar atau lebih panas.
4. Confidence merupakan tingkat kepercayaan deteksi. Kebanyakan bernilai n (normal), tapi ada juga l (low) dan h (high). Ini berguna untuk menyaring data yang paling akurat.

4.2 Kegunaan Fungsi

4.2.1 Data Wrangling Functions

```
#' @return Data frame hotspot yang sudah diproses
load_and_clean_hotspot <- function(csv_path, frp_threshold = 0.5,
                                         confidence_min = "low") {
```

Gambar 2. Fungsi Load and Clean Hotspot

Fungsi ini melakukan validasi, filtering berdasarkan FRP & confidence, bersihkan data NA, mengolah waktu, dan mengembalikan data hotspot yang sudah siap dianalisis.

```
#' @return SF object
convert_to_spatial <- function(hotspot_data, crs = 4326) {
```

Gambar 3. Fungsi Convert to Spatial

Fungsi ini digunakan untuk mengubah data hotspot yang berupa data frame biasa menjadi data spasial (spatial object) menggunakan paket sf.

```
#' @return List dengan hotspot_sf dan regional summary
join_with_shapefile <- function(hotspot_sf, shp_path = NULL) {
```

Gambar 4. Fungsi Join with Shapefile

Fungsi ini digunakan untuk menggabungkan data hotspot yang berbentuk sf/spatial dengan shapefile wilayah agar setiap titik hotspot mengetahui berada di wilayah mana, menyamakan sistem koordinat dengan hotspot, mengembalikan hotspot yang sudah memiliki informasi wilayah.

```
#' @return Data frame dengan grid summary
create_spatial_grid <- function(hotspot_sf, grid_size = 0.1) {
```

Gambar 5. Fungsi Create Spatial Grid

Fungsi ini digunakan untuk membuat grid kotak-kotak berdasarkan latitude & longitude, Menempatkan setiap hotspot ke dalam gridnya, menghitung statistik pada setiap grid, mengembalikan tabel ringkasan grid.

```
#' @return List dengan statistik ringkasan
get_data_summary <- function(hotspot_data) {
```

Gambar 6. Fungsi Get Data Summary

Fungsi ini digunakan untuk membuat ringkasan statistik (summary) dari data hotspot. Ringkasan ini mencakup jumlah data, rentang tanggal, rentang koordinat, statistik FRP, statistik brightness, distribusi confidence, distribusi siang/malam, daftar satelit yang digunakan. Fungsi ini sangat membantu untuk eksplorasi awal dataset (EDA).

4.2.2 Temporal Analysis Functions

```
#' @return Data frame dengan ringkasan bulanan
analyze_monthly_distribution <- function(hotspot_data) {
```

Gambar 7. Fungsi Analyze Monthly Distribution

Fungsi ini digunakan untuk menganalisis distribusi hotspot per bulan, menghitung berbagai statistik bulanan, dan menemukan bulan dengan hotspot terbanyak.

```
#' @return Data frame dengan ringkasan per jam
analyze_hourly_pattern <- function(hotspot_data) {
```

Gambar 8. Fungsi Analyze Hourly Pattern

Fungsi ini digunakan untuk mengelompokkan hotspot berdasarkan jam, menghitung jumlah hotspot per jam, menghitung brightness dan FRP rata-rata per jam, menentukan jam puncak hotspot (peak hour), serta mengembalikan ringkasan statistik per jam.

4.2.3 Clustering Functions

```
#' @return List dengan hasil clustering
perform_kmeans_clustering <- function(hotspot_sf, k = 3, nstart = 25) {
```

Gambar 9. Fungsi Perform K-Means Clustering

Fungsi ini digunakan untuk mengelompokkan hotspot ke dalam beberapa cluster geografis menggunakan algoritma K-Means, dan menghasilkan statistik tiap cluster. Tujuannya untuk mengidentifikasi zona hotspot terpadat, melihat pusat cluster geografis kebakaran, memahami karakteristik hotspot per cluster, membuat analisis spasial berbasis pengelompokan.

```
#' @return List dengan WSS values dan suggested k
find_optimal_k <- function(hotspot_sf, max_k = 10) {
```

Gambar 10. Fungsi Find Optimal K

Fungsi ini digunakan untuk menentukan jumlah cluster (k) terbaik untuk K-Means dengan menggunakan metode Elbow (Elbow Method). Metode ini sangat penting sebelum menjalankan clustering supaya hasil cluster lebih akurat dan tidak asal memilih jumlah cluster.

4.2.4 Bootstrap Validation Functions

```
#' @return List dengan metrik validasi
bootstrap_cluster_validation <- function(hotspot_sf, k = 3, n_boot = 100) {
```

Gambar 11. Fungsi Bootstrap Cluster Validation

Fungsi ini digunakan untuk memvalidasi kestabilan hasil K-Means clustering menggunakan metode bootstrap. Dengan kata lain, fungsi ini mengecek seberapa sensitif hasil cluster terhadap variasi data, sehingga kita bisa menilai apakah cluster stabil / konsisten dan apakah cluster mudah berubah jika data sedikit diubah. Fungsi ini penting untuk mengukur keandalan cluster hotspot geografis.

4.2.5 Validation Functions

```
#' @return Leaflet map object
create_interactive_map <- function(clustering_result) {
```

Gambar 12. Fungsi Create Interactive Map

Fungsi ini digunakan untuk membuat peta interaktif (leaflet) dari hasil clustering hotspot, fungsinya untuk menampilkan titik hotspot di peta dengan warna sesuai cluster, memberikan popup informasi detail tiap hotspot, menambahkan legenda untuk memudahkan interpretasi cluster, memungkinkan eksplorasi hotspot secara visual. Fungsi ini sangat berguna untuk analisis spasial dan presentasi data hotspot.

```
#' @return ggplot object
plot_monthly_trend <- function(monthly_data) {
```

Gambar 13. Fungsi Plot Monthly Trend

Fungsi ini digunakan untuk membuat grafik trend bulanan jumlah hotspot kebakaran hutan dan lahan (karhutla). Tujuannya untuk menampilkan jumlah hotspot per bulan dalam bentuk garis (line chart), menyoroti rata-rata jumlah hotspot per bulan, memudahkan analisis pola musiman atau tren tahunan. Grafik ini berguna untuk monitoring hotspot, laporan, dan visualisasi data temporal.

```
#' @return ggplot object
plot_hourly_pattern <- function(hourly_data) {
```

Gambar 14. Fungsi Plot Hourly Pattern

Fungsi ini digunakan untuk membuat grafik pola deteksi hotspot per jam dalam 24 jam (UTC). Tujuannya untuk Menunjukkan jam-jam aktif hotspot, Membantu melihat trend harian dan waktu puncak deteksi, Memvisualisasikan distribusi hotspot dalam periode 24 jam. Grafik ini berguna untuk analisis temporal harian hotspot.

```
#' @return ggplot object
plot_cluster_spatial <- function(clustering_result) {
```

Gambar 15. Fungsi Plot Cluster Spasial

Fungsi ini digunakan untuk memvisualisasikan hasil K-Means clustering hotspot secara spasial. Fungsinya untuk Menampilkan titik hotspot di peta koordinat geografis (longitude & latitude), Mewarnai titik berdasarkan cluster, Menandai pusat cluster dengan simbol khusus (+), Memberikan konteks visual untuk evaluasi distribusi hotspot. Fungsi ini berguna untuk melihat pola spasial cluster hotspot karhutla.

```
#' @return ggplot object
plot_bootstrap_stability <- function(bootstrap_result) {
```

Gambar 16. Fungsi Plot Bootstrap Stability

Fungsi ini digunakan untuk memvisualisasikan hasil bootstrap validation dari K-Means clustering. Tujuannya untuk Menilai stabilitas clustering melalui distribusi Within-Cluster Sum of Squares (WSS), Menampilkan mean WSS dan interval kepercayaan 95%, Membantu melihat apakah cluster yang terbentuk konsisten dan stabil. Fungsi ini penting untuk validasi kualitas cluster secara statistik.

```
#' @return ggplot object
plot_elbow_method <- function(elbow_result) {
```

Gambar 17. Fungsi Plot Elbow Method

Fungsi ini digunakan untuk Membuat visualisasi metode Elbow untuk menentukan jumlah cluster optimal (K) pada K-Means.

```
#' @return ggplot object
plotRegionalDistribution <- function(regionalData, regionCol, topN = 15) {
```

Gambar 18. Fungsi Plot Regional Distribution

Fungsi ini digunakan untuk memvisualisasikan distribusi hotspot berdasarkan wilayah (regional).

```
#' @return List Lengkap berisi semua hasil analisis
analyze_karhutla <- function(csvPath,
shpPath = NULL,
```

Gambar 19. Fungsi Analyze Karhutla

Fungsi ini digunakan untuk menggabungkan beberapa langkah analisis menjadi satu alur terpadu, mulai dari membuat dan membersihkan data hotspot, konversi data menjadi format spasial (sf), menggabungkan dengan shapefile wilayah, analisis spasial, temporal, dan statistik, K-Means clustering hotspot dan validasi stabilitasnya, visualisasi hasil, menentukan jumlah cluster optimal (opsional), dan menyimpan hasil output.

4.3 Analisis Temporal

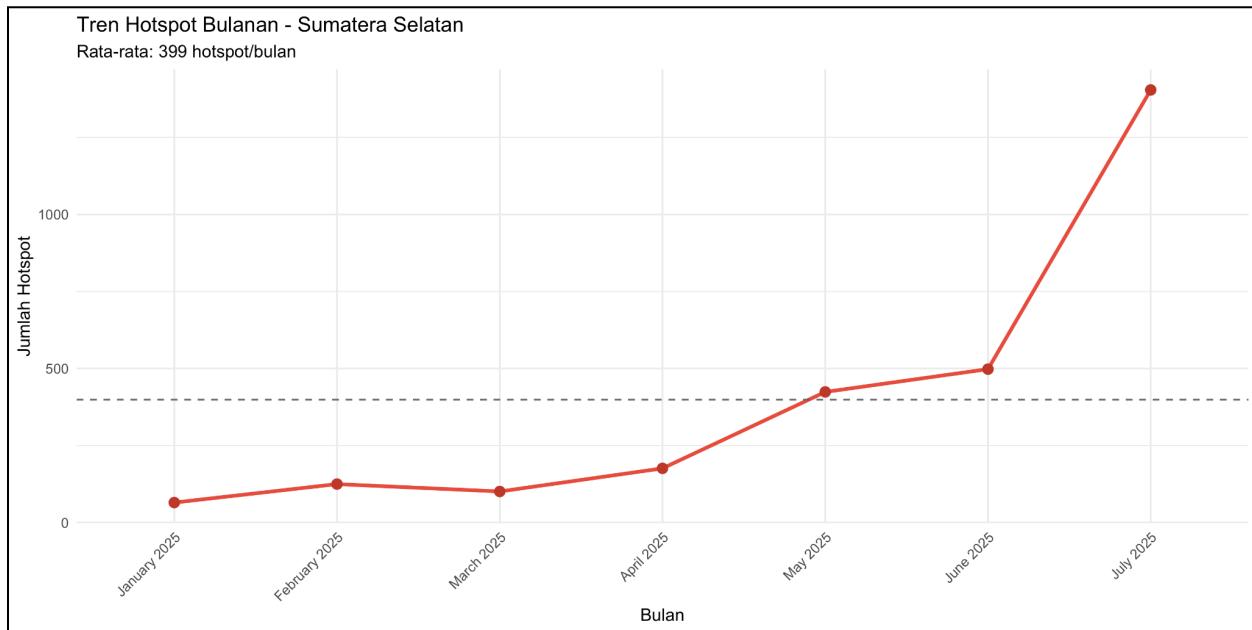
Analisis temporal sangat penting dalam menganalisis karena dapat mempelajari variabel dalam urutan waktu (harian, mingguan, bulanan, tahunan). Dengan tujuan untuk menangkap pola seperti tren, musiman, siklus, dan fluktuasi acak, serta membuat prediksi masa depan.

4.3.1 Analisis Trend Bulanan

Analisis trend bulanan menunjukkan pola musiman yang jelas dalam kejadian hotspot, sebagaimana ditampilkan pada Gambar 19. Grafik menunjukkan peningkatan yang sangat signifikan dari periode Januari hingga Juli 2025.

Tabel 2. Distribusi Hotspot per Bulan

Bulan	Total Hotspot	Avg FRP (MW)	High Confidence
Januari 2025	65	6	0
Februari 2025	125	5.31	0
Maret 2025	101	6.22	0
April 2025	176	5.62	0
Mei 2025	424	6.70	0
Juni 2025	498	6.60	0
Juli 2025	1404	8.29	0
Total	2793	7.46	0



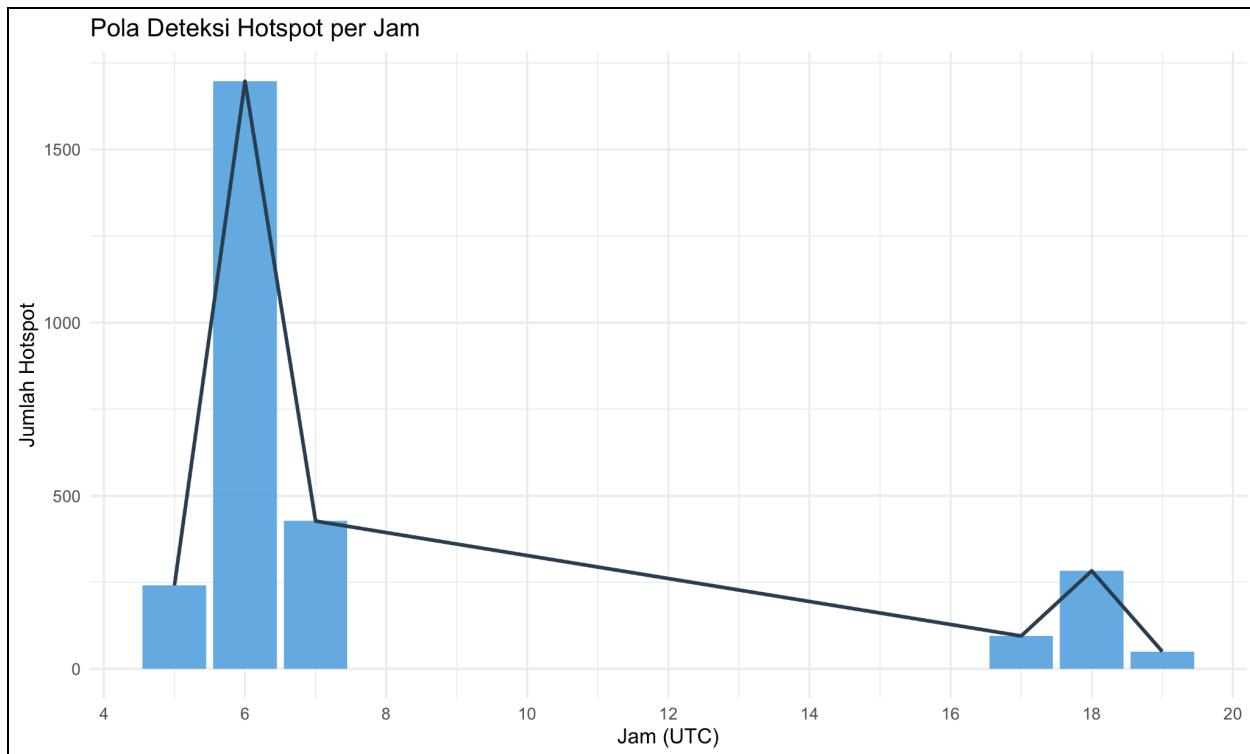
Gambar 20. Tren Hotspot Bulanan - Sumatera Selatan

Gambar 20. memperlihatkan tren peningkatan eksponensial kejadian hotspot dari Januari hingga Juli 2025. Garis horizontal putus-putus menunjukkan rata-rata bulanan sebesar 399 hotspot per bulan. Terdapat peningkatan drastis pada bulan Juli 2025 dengan 1.404 hotspot, yang mencapai 3,5 kali lipat di atas rata-rata dan merepresentasikan 50,2% dari total kejadian selama periode observasi.

Pola ini mengindikasikan adanya musim kemarau yang intensif pada pertengahan tahun 2025. Peningkatan bertahap dari Januari (65 hotspot) hingga April (176 hotspot) menunjukkan transisi dari musim hujan ke musim kemarau. Akselerasi peningkatan mulai terlihat signifikan pada Mei (424 hotspot) dan Juni (498 hotspot), sebelum mencapai puncaknya di Juli. Intensitas api rata-rata (Fire Radiative Power) juga menunjukkan peningkatan dari 5,31 MW di Februari menjadi 8,29 MW di Juli, mengindikasikan tidak hanya frekuensi tetapi juga keparahan kebakaran yang meningkat.

4.3.2 Pola Deteksi Per Jam

Analisis Deteksi Hotspot pada Gambar 20. memperlihatkan pola waktu deteksi hotspot yang sangat khas dengan konsentrasi ekstrem pada periode tertentu.



Gambar 21. Pola Deteksi Hotspot Per Jam

Gambar 21. menampilkan distribusi temporal hotspot berdasarkan jam deteksi dalam waktu UTC. Terlihat jelas adanya puncak yang sangat dominan pada jam 06:00 UTC (setara 13:00 WIB) dengan lebih dari 1.600 deteksi hotspot. Pola ini membentuk distribusi bimodal dengan puncak sekunder yang lebih kecil pada jam 17:00-18:00 UTC (00:00-01:00 WIB dini hari). Garis tren overlay menunjukkan penurunan bertahap dari puncak pagi menuju siang hari, dengan aktivitas minimal pada jam 08:00-16:00 UTC.

Tabel 3. Distribusi Hotspot per Jam (UTC)

Jam (UTC)	Total Hotspot	Avg FRP (MW)	Persentase
05:00-07:00	1697	7.52	60.8%
17:00-19:00	284	7.12	10.2%
Jam lainnya	812	7.35	29%

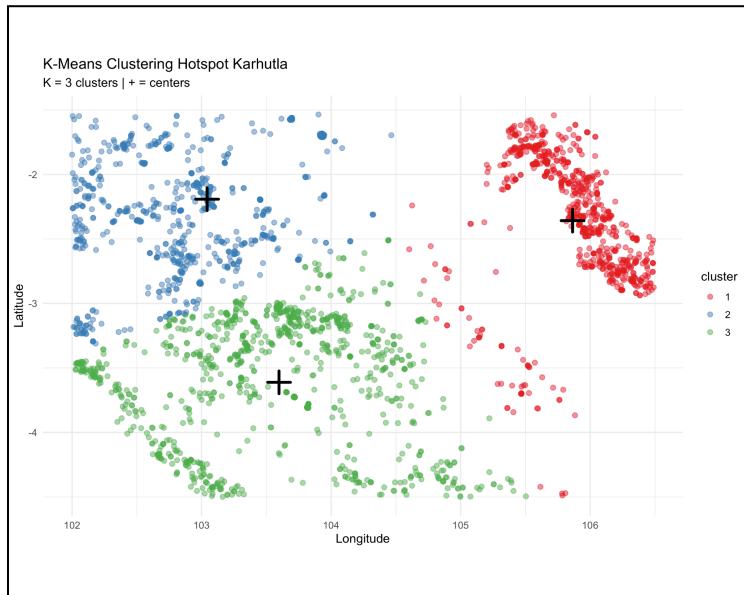
Konsentrasi deteksi pada jam 06:00 UTC mencerminkan karakteristik orbit satelit NOAA-20 yang melintas di wilayah Sumatera Selatan pada waktu tersebut. Pola ini juga berkorelasi dengan kondisi atmosfer yang optimal untuk deteksi thermal pada siang hari ketika kontras suhu antara api dan lingkungan mencapai maksimum. Puncak sekunder pada sore-malam hari mengindikasikan adanya pass satelit kedua, meskipun dengan cakupan yang lebih terbatas.

Pemahaman pola temporal ini penting untuk optimalisasi jadwal monitoring dan respons tim pemadam kebakaran.

4.4 Hasil K-Means Clustering

4.4.1 Statistik Clusterin

K-Means clustering dengan k=3 berhasil mengelompokkan hotspot menjadi tiga zona geografis yang distinct, sebagaimana ditampilkan pada **Gambar 22**.



Gambar 22. K-Means Clustering Hotspot Karhutla

Gambar 22. memvisualisasikan hasil clustering spasial hotspot dengan tiga kelompok berbeda yang ditandai dengan warna merah (Cluster 1), biru (Cluster 2), dan hijau (Cluster 3). Simbol "+" hitam menunjukkan lokasi centroid masing-masing cluster. Pola spatial menunjukkan pemisahan geografis yang jelas: Cluster 1 terkonsentrasi di wilayah timur (longitude $>105^\circ$), Cluster 2 di wilayah barat (longitude $<103.5^\circ$), dan Cluster 3 mendominasi wilayah tengah hingga selatan dengan sebaran yang paling luas.

Metrik Kualitas Clustering:

- Between-cluster SS / Total SS: 74.5%
- Total Within-cluster SS: 1.617,9

Nilai Between-cluster SS sebesar 74,5% mengindikasikan bahwa model mampu menjelaskan hampir tiga perempat variabilitas data melalui pemisahan antar cluster. Ini merupakan indikator yang sangat baik untuk kualitas clustering, menunjukkan bahwa ketiga cluster memiliki

karakteristik spasial yang distinct dan well-separated. Nilai Within-cluster SS yang relatif rendah menunjukkan bahwa titik-titik dalam setiap cluster memiliki kedekatan geografis yang tinggi.

4.4.2 Karakteristik Geografis Cluster

Cluster 1 - Zona Timur (Pantai Timur) mencakup 722 hotspot (25,9%) yang terkonsentrasi di wilayah pesisir Sumatera Selatan bagian timur dengan pusat pada koordinat $106,02^\circ$ BT dan $-2,36^\circ$ LS. Cluster ini menunjukkan nilai Fire Radiative Power tertinggi dengan rata-rata 8,73 MW, mengindikasikan intensitas kebakaran yang lebih parah dibandingkan zona lainnya. Karakteristik geografis wilayah ini yang didominasi lahan gambut dan perkebunan kelapa sawit kemungkinan menjadi faktor pendorong tingginya intensitas kebakaran. Lahan gambut yang mengalami degradasi dan pengeringan sangat rentan terhadap kebakaran dengan intensitas tinggi yang sulit dipadamkan.

Cluster 2 - Zona Barat (Perbatasan Bengkulu) terdiri dari 894 hotspot (32%) yang tersebar di wilayah perbatasan dengan Provinsi Bengkulu pada koordinat pusat $103,03^\circ$ BT dan $-2,19^\circ$ LS. Wilayah ini merupakan area transisi antara dataran rendah dan perbukitan Bukit Barisan dengan topografi yang lebih bergelombang. Rata-rata FRP sebesar 6,87 MW menunjukkan intensitas sedang, kemungkinan karena kombinasi antara lahan pertanian, perkebunan karet, dan sisa hutan sekunder. Jumlah hotspot yang signifikan di zona ini mengindikasikan tekanan pembukaan lahan yang tinggi.

Cluster 3 - Zona Tengah-Selatan (Dominan) merupakan cluster terbesar dengan 1.177 hotspot (42,1%) yang mencakup wilayah luas di bagian tengah hingga selatan Sumatera Selatan, berpusat pada $104,01^\circ$ BT dan $-3,61^\circ$ LS. Meskipun memiliki intensitas rata-rata terendah (6,79 MW), frekuensi kejadian yang sangat tinggi menjadikan zona ini sebagai area prioritas utama untuk monitoring. Wilayah ini mencakup daerah sekitar Palembang dan kabupaten-kabupaten di sekitarnya dengan mixed land use antara pertanian, perkebunan, dan pemukiman. Tingginya frekuensi kejadian kemungkinan terkait dengan praktik pembakaran lahan pertanian dan aktivitas antropogenik lainnya.

4.5 Hasil Bootstrap Validation

4.5.1 Analisis Stabilitas

Bootstrap validation dilakukan dengan 100 iterasi untuk menguji konsistensi hasil clustering.

Tabel 5. Hasil Bootstrap Validation

Metrik	Nilai
Mean WSS	1.620,8
Std Dev WSS	29,2

Metrik	Nilai
Coefficient of Variation (CV)	1.80%
Klasifikasi Stabilitas	Sangat Stabil

Distribusi WSS Bootstrap:

- Range: 1.550 - 1.700
- Distribusi mendekati normal dengan sedikit right-skew
- Tight distribution mengindikasikan hasil yang konsisten

4.5.2 Interpretasi Stabilitas

CV = 1.80% (< 5%) mengindikasikan bahwa:

1. Hasil clustering sangat stabil dan tidak sensitif terhadap variasi sampling
2. Struktur cluster yang ditemukan robust dan dapat diandalkan
3. Pengelompokan geografis konsisten dan bukan artifak dari data tertentu
4. Model dapat digunakan untuk prediksi dan perencanaan dengan confidence tinggi

4.6 Analisis Regional

4.6.1 Top 10 Wilayah dengan Hotspot Terbanyak

Analisis berdasarkan koordinat latitude menunjukkan konsentrasi hotspot di lokasi tertentu. Meskipun analisis ini terbatas karena belum ter-mapping ke nama administratif kabupaten/kota, pola geografis tetap dapat diidentifikasi.

Tabel 6. Distribusi Hotspot per Koordinat Latitude (Top 10)

Rank	Latitude	Total Hotspot	Avg FRP (MW)	Max FRP (MW)
1	-4.49	2	4.03	4.47
2	-3.85	2	5.46	6.59
3	-3.81	2	3.26	4.37
4	-3.80	2	2.72	3.44
5	-3.73	2	48.60	73.70
6	-3.72	2	3.95	4.12

Rank	Latitude	Total Hotspot	Avg FRP (MW)	Max FRP (MW)
7	-3.72	2	4.00	4.07
8	-3.72	2	8.49	9.20
9	-3.72	2	3.34	3.72
10	-3.72	2	2.61	2.78

Analisis Spasial pada data ini dapat memetakan latitude dan longitude dan dapat melihat sebaran spasial titik panas dan mengidentifikasi daerah rawan kebakaran. Intensitas Kebakaran dapat dilihat dari kolom frp yang memungkinkan untuk tidak hanya menghitung berapa banyak titik api, tetapi juga seberapa kuat kebakaran tersebut. Dapat dibandingkan intensitas kebakaran siang vs malam (daynight) atau antar wilayah yang berbeda. Validasi Data dapat dilakukan pada kolom confidence dan bisa juga digunakan untuk memfilter data.

BAB IV

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pengembangan fungsi R dan analisis data hotspot VIIRS di Provinsi Sumatera Selatan periode Januari hingga Juli 2025, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil mengembangkan fungsi terpadu dalam bahasa pemrograman R yang mampu mengotomatisasi seluruh alur kerja (workflow) analisis komputasi statistik, mulai dari data *wrangling* (integrasi data CSV dan Shapefile), ekstraksi fitur temporal, hingga konversi data spasial secara efisien.
2. Melalui teknik *feature engineering* dan agregasi temporal berbasis komputasi, fungsi yang dikembangkan mampu mendeteksi pola *diurnal* (harian) dan musiman yang tersembunyi dalam *dataset*. Algoritma berhasil memetakan konsentrasi puncak deteksi pada pukul 06:00 UTC (13:00 WIB) dan mengidentifikasi tren peningkatan eksponensial hotspot yang memuncak pada bulan Juli 2025 (1.404 titik).
3. Implementasi algoritma K-Means Clustering ($k=3$) berhasil mencapai konvergensi yang meminimalkan variansi dalam klaster (*Within-Cluster Sum of Squares*). Secara komputasional, algoritma ini mengidentifikasi tiga zona karakteristik karhutla. Cluster 1 (Zona Timur/Pantai Timur) teridentifikasi sebagai wilayah dengan intensitas kebakaran tertinggi (rata-rata FRP 8,73 MW), yang diindikasikan sebagai area lahan gambut yang rentan. Sementara itu, Cluster 3 (Zona Tengah-Selatan) memiliki frekuensi kejadian terbanyak (42,1% dari total data).
4. Penerapan metode resampling bootstrap dengan 100 iterasi membuktikan bahwa hasil klasterisasi sangat stabil dan dapat diandalkan. Hal ini ditunjukkan oleh nilai Coefficient of Variation (CV) sebesar 1,80%, yang mengindikasikan bahwa pembagian zona wilayah tersebut konsisten dan bukan merupakan artefak statistik semata.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan selama pengembangan fungsi, penulis mengajukan beberapa saran untuk pengembangan selanjutnya dan pemangku kebijakan:

1. Saran untuk Pengembangan Fungsi R:
 - a. Disarankan untuk menambahkan fitur Spatial Join yang lebih mendalam pada fungsi R agar tabel luaran tidak hanya menampilkan koordinat latitude/longitude,

tetapi secara otomatis menyertakan nama administratif (Kabupaten/Kecamatan) untuk mempermudah interpretasi pengguna.

- b. Pengembangan selanjutnya dapat membandingkan metode K-Means dengan algoritma berbasis densitas seperti DBSCAN untuk menangani noise atau titik panas yang tidak membentuk klaster membulat.
2. Saran untuk Kebijakan Mitigasi:
- a. Mengingat Cluster 1 (Zona Timur) memiliki intensitas api (FRP) tertinggi, pemerintah daerah disarankan untuk memprioritaskan alokasi sumber daya pemadaman berat (seperti water bombing) di wilayah pesisir timur, terutama pada lahan gambut.
 - b. Patroli dan pemantauan lapangan disarankan untuk diintensifkan pada rentang waktu 12:00 – 14:00 WIB, mengingat pola deteksi satelit menunjukkan puncak kemunculan titik panas pada jam tersebut.
 - c. Upaya pencegahan dini harus dimaksimalkan sebelum bulan Mei, mengingat tren data menunjukkan akselerasi peningkatan hotspot yang signifikan dimulai pada bulan tersebut sebelum memuncak di bulan Juli.

DAFTAR PUSTAKA

- [1] A. Zaskia, B. H. Saharjo, and I. Albar, "Pengaruh curah hujan terhadap kejadian kebakaran hutan dan lahan di Kabupaten Banyuasin Provinsi Sumatera Selatan," *J. Silvikultur Tropika*, vol. 14, no. 3, pp. 195–200, Dec. 2023, doi: 10.29244/j-siltrop.14.03.195-200.
- [2] H. S. D. Kospa and A. A. Praja, "Evaluasi pencegahan kebakaran hutan dan lahan gambut di Kabupaten Ogan Komering Ilir Sumatera Selatan," *J. Rekayasa*, vol. 13, no. 1, pp. 1–9, Jun. 2023, doi: 10.37037/jrftsp.v13i1.165.
- [3] B. H. Saharjo and M. R. A. Nasution, "Pola sebaran titik panas (hotspot) sebagai indikator terjadinya kebakaran hutan dan lahan di Kabupaten Aceh Barat," *J. Silvikultur Tropika*, vol. 12, no. 2, pp. 60–66, Aug. 2021, doi: 10.29244/j-siltrop.12.2.60-66.
- [4] P. Bruce, A. Bruce, and P. Gedeck, *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2020.R.
- [5] Lovelace, J. Nowosad, and J. Muenchow, *Geocomputation with R*, 1st ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2019, doi: 10.1201/9780203730058.
- [6] H. Wickham, M. Çetinkaya-Rundel, and G. Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2023.
- [7] M. L. Rizzo, *Statistical Computing with R*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall/CRC, 2019.
- [8] F. Li, X. Zhang, S. Kondragunta, and I. Csiszar, "Comparison of fire radiative power estimates from VIIRS and MODIS observations," *J. Geophys. Res. Atmos.*, vol. 123, no. 22, pp. 12 725–12 742, Nov. 2018, doi: 10.1029/2017JD027823.

LAMPIRAN

A. Kode Program

Akses Kode Lengkap: [drive-kode-program](#)

```

#> Fungsi Utama: Analisis Karhutla Lengkap
#> @param csv_path Path ke file CSV hotspot VIIRS
#> @param shp_path Path ke shapefile (optional), NULL jika tidak ada
#> @param k Jumlah cluster untuk K-Means (default: 3)
#> @param n_boot Jumlah iterasi bootstrap (default: 100)
#> @param frp_threshold FRP minimum w (default: 0.5)
#> @param confidence_min Confidence min ("low", "medium", "high")
#> @param find_k_optimal Apakah mencari k optimal dulu? (default: FALSE)
#> @param max_k Maksimum k untuk elbow method (default: 10)
#> @param output_dir Direktori output (default: "hasil_karhutla")
#> @return List lengkap berisi semua hasil analisis
analyze_karhutla <- function(csv_path,
                               shp_path = NULL,
                               k = 3,
                               n_boot = 100,
                               frp_threshold = 0.5,
                               confidence_min = "low",
                               find_k_optimal = FALSE,
                               max_k = 10,
                               output_dir = "hasil_karhutla") {
  cat("\n")
  =====
  cat(" ANALISIS HOTSPOT KARHUTLA SUMATERA SELATAN\n")
  cat(" Pengembangan Fungsi R untuk Data VIIRS FIRMS\n")
  cat("=====\n")
  start_time <- Sys.time()

  # ===== STEP 1: DATA WRANGLING =====
  cat("[STEP 1/7] DATA WRANGLING\n")
  cat("----\n")
  hotspot <- load_and_clean_hotspot(csv_path, frp_threshold, confidence_min)
  hotspot_sf <- convert_to_spatial(hotspot)
  spatial_result <- join_with_shapefile(hotspot_sf, shp_path)
}
  
```

Gambar 23. Dokumentasi Kode Program

B. Sample Data Mentah

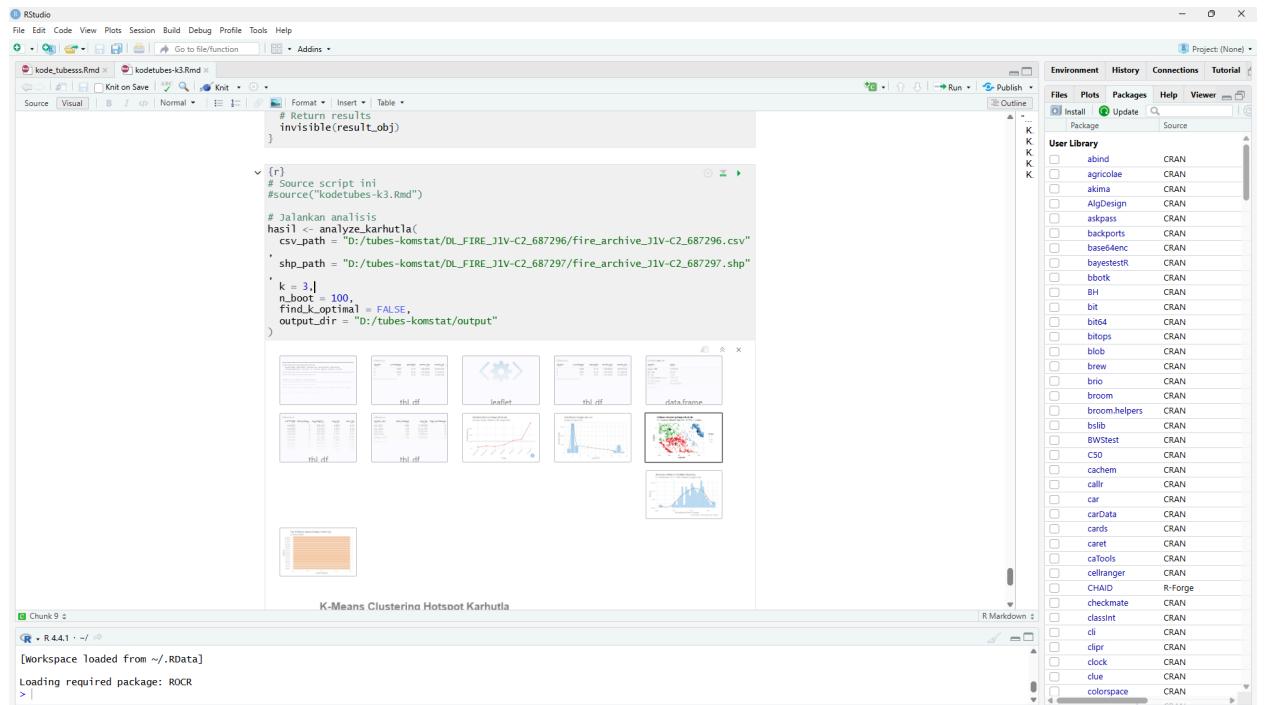
Akses data selengkapnya dapat dilihat pada tautan berikut : [data-hotspot-sumsel](#)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	latitude	longitude	brightness	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_l31	frp	daynight	type
2	-4.39075	103.12169	340.72	0.38	0.59	2025-01-02	556 N20	VIIRS	n	2	200.82	2.9	D	0	
3	-4.39749	103.12086	342.96	0.38	0.59	2025-01-02	556 N20	VIIRS	n	2	291.4	3.76	D	0	
4	-3.7382	103.08326	349.23	0.38	0.58	2025-01-02	557 N20	VIIRS	n	2	286.59	7.6	D	0	
5	-3.73865	103.08002	332.28	0.38	0.59	2025-01-02	557 N20	VIIRS	n	2	286.96	6.37	D	0	
6	-2.19342	103.45049	311.68	0.45	0.39	2025-01-02	1839 N20	VIIRS	n	2	283.57	2.07	N	0	
7	-2.19646	103.45075	305.22	0.45	0.39	2025-01-02	1839 N20	VIIRS	n	2	284.18	1.48	N	0	
8	-3.31532	104.32404	301.86	0.39	0.36	2025-01-03	1820 N20	VIIRS	n	2	281.28	0.37	N	0	
9	-3.06842	104.20195	305.88	0.39	0.36	2025-01-03	1820 N20	VIIRS	n	2	276.56	0.8	N	0	
10	-4.0678	102.64227	339.39	0.47	0.48	2025-01-04	659 N20	VIIRS	n	2	287.27	8.69	D	0	
11	-2.84222	102.8253	340.59	0.5	0.49	2025-01-04	700 N20	VIIRS	n	2	280.17	6.9	D	0	
12	-3.66599	102.35732	334.63	0.45	0.47	2025-01-04	700 N20	VIIRS	n	2	291.38	16.28	D	0	
13	-3.28095	104.36423	345.01	0.63	0.54	2025-01-04	700 N20	VIIRS	n	2	292.4	13.88	D	0	
14	-3.17682	104.06651	344.42	0.6	0.53	2025-01-04	700 N20	VIIRS	n	2	287.39	14.46	D	0	
15	-4.44002	103.23213	332.03	0.41	0.37	2025-01-05	640 N20	VIIRS	n	2	293.61	1.71	D	0	
16	-4.23124	102.88432	345.13	0.41	0.37	2025-01-05	640 N20	VIIRS	n	2	298.2	8.52	D	0	
17	-4.45353	102.96379	332.26	0.41	0.37	2025-01-05	640 N20	VIIRS	n	2	296.38	1.49	D	0	
18	-1.96727	102.18089	335.55	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	292.22	8.52	D	0	
19	-1.92735	102.06854	329.15	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	292.86	1.56	D	0	
20	-3.46725	102.09218	330.52	0.39	0.36	2025-01-05	641 N20	VIIRS	n	2	296.71	2.85	D	0	
21	-4.05852	102.75307	332.22	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	294.47	2.83	D	0	
22	-4.14113	102.67478	340.93	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	296.27	16.06	D	0	
23	-2.03781	102.20372	325.12	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	294.22	2.21	D	0	
24	-3.31142	102.18249	333.73	0.39	0.37	2025-01-05	641 N20	VIIRS	n	2	291.64	5.28	D	0	
25	-3.78477	102.41368	334.89	0.4	0.37	2025-01-05	641 N20	VIIRS	n	2	295.67	3.81	D	0	
26	-3.19036	102.03681	329.48	0.6	0.7	2025-01-08	725 N20	VIIRS	n	2	283.11	3.97	D	0	
27	-3.19072	102.03342	337.5	0.6	0.7	2025-01-08	725 N20	VIIRS	n	2	283.46	4.9	D	0	
28	1.59362	103.42684	319.81	0.38	0.36	2025-01-08	1920 N20	VIIRS	n	2	290.43	4.24	N	0	

Gambar 24. Dokumentasi Kode Program

C. Dokumentasi Hasil Pengembangan Fungsi R

Akses hasil penggunaan fungsi dapat dilihat pada link berikut : [Hasil Pengembangan](#)



Gambar 25. Dokumentasi Kode Program