Pengelompokan Data Hotspot Menggunakan Metode LOF K-Means

Pambudi, O.T.[1], Kurniawan, M.[2], Muhima, R.R.[3], Hakimah, M.[4]
Institut Teknologi Adhi Tama Surabaya[1,2,3,4]
e-mail:oktavianpilu@gmail.com

ABSTRACT

Forest fires become one of the problems in Indonesia.From beginning of the year to September in 2019, forest and land fires in Indonesia reached 857.756 hectares.The impactof forest fires can disrupt health, transportation and even bilateral relations with neighboring countries. Hotspots are indications of a forest or land fire. Clustering of hotspot as a way of analyzing data is crucial to prevent forest fires or land from becoming larger. In previous research, clustering of hotspot data based latitude and longtitude feature, with LOF K-Means gave better results than K-Means. This research conducted a data hotspot grouping with the LOF K-Means method based on latitude, longtitude and brightness features. The purpose of this research is to know the results of comparisons between the data hotspots based on latitude, longtitude with hotspot data clustering based on latitude, longtitude and brightness features. The results of this research became the foundation for subsequent research as the prevention of losses caused by forest fires and larger land. Based on the results of the research, the addition of brightness features as the basis of clustering using the LOF K-Means method does not give better results.

Kata kunci: LOF K-Means, *hotspot, feature, clustering*

ABSTRAK

Kebakaran hutan menjadi salah satu permasalahan di Indonesia. Dari awal tahun hingga September pada tahun 2019, kebakaran hutan dan lahan di Indonesia mencapai 857.756 hektar. Dampak kebakaran hutan dapat mengganggu kesehatan, transportasi bahkan hubungan bilateral dengan negara tetangga. Titik panas bumi atau *hotspot* merupakan indikasi adanya kebakaran hutan atau lahan. Pengelompokan data *hotspot* sebagai cara analisa data *hotspot* sangat penting untuk mencegah terjadinya kebakaran hutan atau lahan menjadi lebih besar. Pada penelitian sebelumnya, pengelompokan data *hotspot* berdasarkan fitur *latitude, longtitude* dengan LOF K-Means memberikan hasil lebih baik dibandingkan K-Means. Penelitian ini dilakukan pengelompokan data *hotspot* dengan metode LOF K-Means berdasarkan fitur *latitude*, *longtitude* dan *brightness*. Tujuan penelitian ini untuk mengetahui hasil perbandingan antara pengelompokan data *hotspot* berdasarkan fitur *latitude, longtitude* dengan pengelompokan data *hotspot* berdasarkan fitur *latitude, longtitude* dan *brightness*. Hasil penelitian ini dijadikan landasan untuk penelitian selanjutnya sebagai upaya pencegahan kerugian akibat kebakaran hutan dan lahan yang lebih besar. Berdasarkan hasil penelitian, penambahan fitur *brightness* sebagai dasar pengelompokan menggunakan metode LOF K-Means tidak memberikan hasil yang lebih baik.

Kata kunci: LOF K-Means, *hotspot*, fitur, pengelompokan

PENDAHULUAN

Salah satu permasalahan Indonesia adalah kebakaran hutan. Dari awal tahun 2019 hingga September 2019, kebakaran hutan dan lahan di Indonesia mencapai 857.756 hektar[1]. Salah satu akibat terjadinya kebakaran hutan maupun lahan adalah kabut asap. Kabut asap dengan jumlah besar dapat mengganggu kesehatan, transportasi, bidang pertanian bahkan hubungan bilateral dengan negara tetangga[2].

Titik panas bumi atau sering disebut *hotspot* merupakan indikasi adanya kebakaran hutan atau lahan[3]. Analisa data *hotspot* sangat penting untuk mencegah terjadinya kebakaran hutan atau lahan menjadi lebih besar[4]. Alat fundamental untuk analisis data adalah pengelompokan atau *clustering* [5]. *Clustering* digunakan untuk mengelompokkan sekumpulan data menjadi beberapa kelompok sehingga data dalam satu kelompok memiliki kesamaan dan berbeda dengan data di kelompok lain [6].

K-Means merupakan salah satu metode *clustering* yang sensitif terhadap pencilan. Pada penelitian sebelumnya [7], metode LOF K-Means yaitu penambahan metode *Local Outlier Factor* (LOF) pada metode K-Means digunakan untuk mengelompokan data *hotspot*. LOF K-Means memberikan hasil lebih baik daripada metode K-Means. Pengelompokan data *hotspot* pada penelitian tersebut didasarkan pada fitur *latitude* dan *longtitude.*

Fitur pada data *hotspot* yang diambil dari web resmi NASA terdapat 11 fitur yaitu latitude, longitude, bright_ti4, scan, track, acq_date, acq_time, satellite, confidence, version, bright_ti5, frp, daynight. Fitur brightness merupakan piksel kecerahan api dan dijadikan dasar pengelompokan pada penelitian [8].

Berdasarkan latar belakang di atas, penelitian ini dilakukan untuk mengelompokan data *hotspot* dengan metode LOF K-Means berdasarkan fitur *latitude, longtitude* dan *brightness*. Dengan penambahan fitur *brightness* sebagai dasar pengelompokan, penelitian ini bertujuan untuk mengetahui hasil perbandingan antara pengelompokan data *hotspot* berdasarkan *latitude, longtitude* dengan pengelompokan data *hotspot* berdasarkan fitur *latitude, longtitude* dan *brightness*. Hasil penelitian ini dijadikan landasan untuk penelitian selanjutnya sebagai upaya pencegahan kerugian akibat kebakaran hutan dan lahan yang lebih besar.

TINJAUAN PUSTAKA

Titik Panas (*Hotspot*)

Secara definisi, titik panas atau *hotspot* merupakan daerah dengan suhu permukaan relatif lebih tinggi dibandingkan daerah di sekitarnya berdasarkan ambang batas suhu tertentu yang terpantau oleh satelit penginderaan jauh[9].

Menurut [9], titik panas atau *hotspot* adalah parameter yang diturunkan dari data satelit dan dindikasikan sebagai lokasi kebakaran hutan dan lahan. Berdasarkan data *hotspot* yang diambil dari https://firms.modaps.eosdis.nasa.gov/active_fire, terdapat 11 fitur yaitu latitude, longitude, bright_ti4, scan, track, acq_time, acq_date, satellite, confidence, version, bright_ti5, frp, daynight.

Clustering

Salah satu teknik data mining adalah algoritma *clustering*[10]. *Clustering* merupakan suatu proses pengelompokan suatu data menjadi beberapa kelompok atau klaster sehingga data dalam satu klaster memiliki kesamaan atau kemiripan dan data tersebut memiliki perbedaan dengan data pada klaster yang lain[11],[12]. Metode *clustering* sendiri terbagi menjadi dua, yaitu *partitional clustering* dan *hierarchial clustering*.

K-Means

K-Means merupakan salah satu metode *clustering* yang termasuk *partitional clustering*. Tujuan dari metode ini adalah memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster [11]. Langkah-langkah metode K-Means untuk mengelompokan sejumlah data *D* dengan *k* (jumlah klaster) yang ditentukan di awal, yaitu[11],[13]:

(1) Menentukan pusat klaster (c_j) sejumlah *k*, biasanya dibangkitkan secara acak.
(2) Menghitung jarak tiap data (x_j) dengan pusat klaster (c_j) dengan persamaan Euclidean

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^{n}(x_j - c_j)^2} \quad \text{.....(1)}$$

(3) Menempatkan data (x_j) ke pusat klaster (c_j) terdekat sehingga data akan berkelompok.
(4) Menentukan pusat klaster baru dengan cara menghitung rata-rata jarak semua data (x_j) dalam satu klaster terhadap pusat klasternya.
(5) Mengulang langkah (2) dan (3) hingga pusat klasternya tidak berubah.

Local Outlier Factor (LOF)

Local Outlier Factor (LOF) merupakan perbandingan kepadatan lokal lingkungan sebuah data dengan kepadatan lokal tetangganya. Nilai LOF ini yang menentukan suatu data dikatakan termasuk pencilan (*outlier*) atau bukan. Suatu data dikatakan pencilan jika nilai LOF nya tinggi yaitu LOF >> 1. Sedangkan jika nilai LOFnya rendah atau mendekati 1, data tersebut bukan pencilan[14]. Nilai LOF suatu titik $a(x_j, y_j)$ dapat dihitung menggunakan persamaan berikut[7],[15]:

$$LOF_k(a) = \frac{\sum_{o \in N_{k-dist(a)}} \frac{lrd\,(o)}{lrd\,(a)}}{|N_{k-dist(a)}|} \quad \ldots(2)$$

$$lrd(a) = \frac{1}{\left[\frac{\sum_{o \in N_{k-dist(a)}} reach-dist_k(a,o)}{|N_{k-dist(a)}|}\right]} \quad \ldots(3)$$

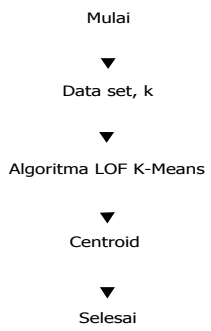
dengan $lrd(a)$ merupakan kepadatan lokal dari suatu objek atau data a; $reach - dist_k(a,o)$ merupakan $max\{k - dist(a), dist(a,o)\}$; sedangkan $N_{k-dist(a)}$ merupakan jumlah tetangga terdekat yang jaraknya tidak melebihi $k - dist(a)$.

$k - dist(a)$ merupakan jarak maksimal dari a terhadap tetangga terdekatnya. $k - dist(a)$ didefinisikan $d(a,o)$ dimana a dan objek $o \in D$ memiliki:

(1) Untuk setidaknya k objek $o' \in D\backslash\{a\}$ dan dinyatakan bahwa $d(a,o') \leq d(a,o)$
(2) Untuk setiap k-1 objek $o' \in D\backslash\{a\}$ dan dinyatakan bahwa $d(a,o') \leq d(a,o)$

METODE

Secara keseluruhan sistem pengelompokan data *hotspot* menggunakan LOF K-Means digambarkan pada gambar 1. Data *hotspot* yang digunakan pada penelitian ini adalah data wilayah Asia Tenggara selama 7 hari berturut-turut yang di ambil dari (https://firms.modaps.eosdis.nasa.gov/active_fire/). Dataset *hotspot* tersebut terdiri dari 11 fitur dengan jumlah data sebanyak 8979 data. Fitur yang digunakan adalah latitude, longtitude dan brightness (*bright-ti4*). Metode yang digunakan dalam mengelompokan data hotspot adalah LOF K-Means hingga diperoleh nilai centroid masing-masing kelompok yang dijadikan titik pusat pemadaman.



Gambar 1. Diagram Alur Pengelompokan Data *Hotspot* dengan Metode LOF K-Means [7]

Diagram blok LOF K-Means pada penelitian ini ditunjukkan pada gambar 2. Data hotspot awalnya dikelompokkan dengan K-Means berdasarkan fitur latitude dan longtitude. Hasil dari pengelompokan dihitung nilai LOF setiap data di masing-masing klaster berdasarkan fitur latitude

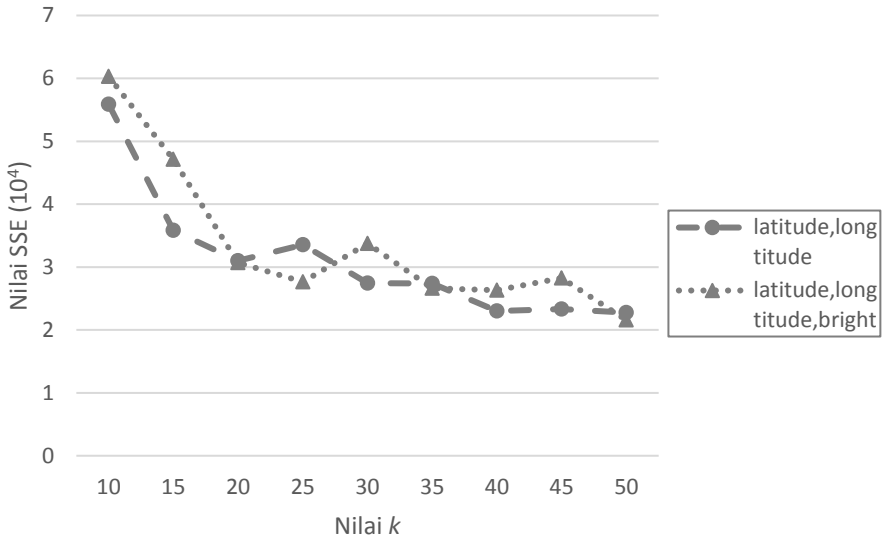
longtitude. Data *hotspot* yang dideteksi pencilan kemudian dihapus. Langkah selanjutnya dihitung nilai LOF dari setiap klaster berdasarkan fitur *brightness*. Data-data yang dideteksi pencilan dihapus. Langkah selanjutnya menentukan menentukan pusat klaster baru menggunakan metode K-Means.



Gambar 2. Diagram Blok Metode LOF K-Means Berdasarkan Fitur *Latitude Longtitude* dan *Brightness*

HASIL DAN PEMBAHASAN

Hasil perbandingan nilai Sum Square Error (SSE) pengelompokan metode LOF K-Means berdasarkan fitur *Latitude Longtitude* dan berdasarkan fitur *Latitude Longtitude, Brightness* ditunjukkan gambar 3. Variasi nilai *k* yang digunakan pada penelitian ini adalah 10, 15, 20, 25, 30,35, 40, 45, dan 50.



Gambar 3. Nilai *Sum Square Error* Metode LOF K-Means Berdasarkan Fitur *Latitude Longtitude* dan Berdasarkan Fitur *Latitude Longtitude, Brightness*

Dari gambar 3, nilai SSE pada k=25 untuk pengelompokkan LOF K-Means berdasarkan fitur *latitude longtitude,brightness* lebih baik secara signifikan dan di k=20,35 dan 50 sedikit lebih baik. Secara umum pengelompokan LOF K-Means berdasarkan fitur *latitude longtitude* secara umum memberikan hasil lebih baik dengan rata-rata $3{,}1133378.10^4$. Sedangkan rata-rata pengelompokan LOF K-Means berdasarkan *latitude longtitude,brightness* $3{,}358356.10^4$.

KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan bahwa pengelompokan dengan metode LOF K-Means berdasarkan fitur *latitude longtitude* memberikan hasil lebih baik daripada berdasarkan fitur *latitude, longtitude, brightness* dengan rata-rata nilai SSE sebesar $3{,}133378.10^4$.

Penambahan fitur *brightness* sebagai dasar pengelompokan menggunakan metode LOF K-Means tidak memberikan hasil yang lebih baik.

DAFTAR PUSTAKA

[1] I. Nugroho, "Kebakaran Hutan dan Lahan Sampai September 2019 Hampir 900 Ribu Hektar," *Mongabai*, 2019. [Online]. Available: https://www.mongabay.co.id/2019/10/22/kebakaran-hutan-dan-lahan-sampai-september-2019-hampir-900-ribu-hektar/.

[2] K. A. Wibowo, "Manajemen Penanganan Kebakaran Hutan dan Lahan (Karhutla) Guna Peningkatan Ekonomi Kerakyatan," *J. Stud. Sos. dan Polit.*, vol. 3, no. 1, pp. 69–83, 2019.

[3] I. Hakiki, A. Ihwan, and J. Sampurno, "Prediksi Kemunculan Titik Panas (Hotspot) Menggunakan Metode Jaringan Syaraf Tiruan Propagasi Balik Studi Kasus di Pontianak," *Prism. Fis.*, vol. III, no. 2, pp. 75–78, 2015.

[4] N. L. Febriana and I. S. Sitanggang, "Outlier Detection on Hotspot Data in Riau Province using OPTICS Algorithm," in *IOP Conference Series: Earth and Environmental Science*, 2017, pp. 1–7.

[5] D. F. Pramesti, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017.

[6] A. Barai and L. Dey, "Outlier Detection and Removal Algorithm in K-Means and Hierarchical Clustering," *World J. Comput. Appl. Technol.*, vol. 5, no. 2, pp. 24–29, 2017.

[7] R. R. Muhima, M. Kurniawan, and O. T. Pambudi, "A LOF K - Means Clustering on Hotspot Data," *Int. J. Artif. Intell. Robot.*, vol. 2, no. 1, pp. 29–33, 2020.

[8] B. K. Amijaya, M. T. Furqon, and C. Dewi, "Clustering Titik Panas Bumi Menggunakan Algoritme Affinity Propagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3835–3842, 2018.

[9] Lapan, "Informasi Titik Panas (Hotspot) Kebakaran Hutan / Lahan," vol. ISBN 978-6, 2016.

[10] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat," *CogITo Smart J.*, vol. 2, no. 2, p. 108, 2016.

[11] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.

[12] U. R. Raval and C. Jani, "Implementing & Improvisation of K-means Clustering Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 55, no. 5, pp. 191–203, 2016.

[13] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k Means Clustering algorithm for prediction of Students Academic Performance," vol. 7, pp. 292–295, 2010.

[14] N. Idham, "Penerapan Outlier Analysis sebagai Salah Satu Rekomendasi Kelompok Belajar Terhadap Siswa Kelas 6 di SDN Pagelaran II," *J. Ilm. Komput. dan Inform.*, 2017.

[15] M. M. Breuniq, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *SIGMOD Rec. (ACM Spec. Interes. Gr. Manag. Data)*, vol. 29, no. 2, pp. 93–104, 2000.

ISSN (print) : 2686-0023
ISSN (online) : 2685-6875