

## Deteksi Anomali Aktivitas Kegempaan Gunung Marapi Menggunakan Algoritma *Local Outlier Factor*

Nopita Sari Murtafiah\*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

\*nopitasarimurtafiah99@gmail.com, nusarhajarisman@unisba.ac.id

**Abstract.** Anomalies in volcano monitoring can occur due to sudden changes in seismic data, ground deformation, gas emissions, or other activities that indicate a potential eruption. Mount Marapi is routinely monitored using seismic, deformation, visual, and geochemical methods, with seismic monitoring being the most commonly used method because seismic activity increases before an eruption. Detecting anomalies in volcanoes is crucial for identifying early signs of volcanic activity that may lead to an eruption. Early detection allows for mitigation measures and evacuation to be carried out, minimizing the impact and loss of life caused by an eruption. One algorithm that can be used to detect anomalies is the Local Outlier Factor (LOF). LOF calculates the density of each data point, where data with significantly lower density is considered an anomaly. This study aims to determine the number and characteristics of data detected as anomalies. In this study, the Local Outlier Factor algorithm will be used to detect anomalies in seismic activity data from Mount Marapi between October 2023 and January 2024. The results of the study show that the LOF algorithm successfully detected 43 anomaly events, or 34.96% of the total data. Data identified as anomalies generally have characteristics such as higher frequency, S-P, amplitude, and duration values than the average. The LOF algorithm successfully detected anomalies on November 29, 2023, approximately four days before the eruption that occurred on December 3, 2023. Additionally, the data detected as anomalies generally have closely related or even consecutive dates or times of occurrence.

**Keywords:** *Anomaly Detection, Mount Marapi, Local Outlier Factor.*

**Abstrak.** Anomali dalam pemantauan gunung api dapat terjadi akibat perubahan mendadak dalam data seismik, deformasi tanah, emisi gas, atau aktivitas lain yang menandakan potensi erupsi. Gunung Marapi dipantau rutin dengan metode seismik, deformasi, visual, dan geokimia, dengan seismik sebagai metode yang paling dominan digunakan karena aktivitas kegempaan meningkat sebelum terjadinya erupsi. Deteksi anomali pada gunung berapi penting dilakukan untuk mengidentifikasi tanda-tanda awal aktivitas vulkanik yang dapat mengarah pada erupsi. Dengan deteksi dini, langkah mitigasi dan evakuasi dapat dilakukan untuk meminimalisir dampak dan korban jiwa yang diakibatkan oleh erupsi. Salah satu algoritma yang dapat digunakan untuk mendeteksi anomali adalah *Local Outlier Factor* (LOF). LOF menghitung kerapatan setiap titik data, di mana data dengan kerapatan jauh lebih rendah dianggap anomali. Penelitian ini bertujuan untuk mengetahui jumlah dan karakteristik data yang terdeteksi anomali. Pada penelitian ini, akan digunakan algoritma *Local Outlier Factor* untuk mendeteksi anomali pada data aktivitas kegempaan Gunung Marapi periode Oktober 2023 hingga Januari 2024. Hasil penelitian menunjukkan bahwa algoritma LOF berhasil mendeteksi 43 kejadian anomali, atau 34,96% dari keseluruhan data. Data yang terdeteksi sebagai anomali umumnya memiliki karakteristik berupa nilai frekuensi, S-P, amplitudo, dan durasi yang lebih tinggi daripada rata-rata. Algoritma LOF berhasil mendeteksi anomali pada 29 November 2023, sekitar empat hari sebelum erupsi yang terjadi pada tanggal 3 Desember 2023. Selain itu, data yang terdeteksi sebagai anomali umumnya juga memiliki tanggal atau waktu kejadian yang berdekatan atau bahkan berurutan.

**Kata Kunci:** *Deteksi Anomali, Gunung Marapi, Local Outlier Factor.*

## A. Pendahuluan

*Data mining* adalah proses ekstraksi atau penggalian informasi yang tersembunyi dari suatu set data besar. Salah satu teknik yang umum digunakan dalam *data mining* adalah klasterisasi, yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok yang serupa berdasarkan pola atau karakteristik tertentu. Pada klasterisasi, *outlier* merupakan data yang tidak termasuk ke dalam suatu klaster dan dianggap sebagai suatu gangguan dan ingin dihilangkan. Meskipun *outlier* biasanya dianggap sebagai gangguan dan ingin dihilangkan, namun keberadaan *outlier* seringkali merupakan informasi yang berharga. *Outlier* bisa saja merupakan representasi dari situasi yang unik atau kejadian langka yang ingin diidentifikasi dan dipelajari lebih lanjut. Oleh karena itu, diperlukan deteksi anomali untuk menemukan pola dalam data yang menyimpang dari perilaku yang diharapkan. Kemampuan mendeteksi anomali tersebut sangat penting dalam berbagai konteks seperti deteksi penipuan pada kartu kredit dan transaksi bank, klaim asuransi, dan pencucian uang (Agustin Nuriani Sirodj et al., 2023). Mendeteksi anomali secara tepat waktu memungkinkan seseorang untuk mengambil tindakan atau membatasi dampak negatif dari anomali tersebut.

*Density-based outlier detection* merupakan salah satu pendekatan deteksi anomali di mana anomali diidentifikasi berdasarkan kerapatan data disekitarnya. *Local Outlier Factor* termasuk ke dalam *Density-based outlier detection*. *Local Outlier Factor* bekerja dengan membandingkan kerapatan (*density*) suatu titik data dengan kerapatan titik-titik data di sekitarnya (tetangganya). Titik data dengan kerapatan jauh lebih rendah dibandingkan tetangganya atau titik data yang memiliki skor *Local Outlier Factor* (LOF) terbesar kemungkinan besar merupakan anomali. Pemilihan algoritma *Local Outlier Factor* didasarkan pada keunggulannya dalam menangani data yang kompleks dan beragam serta tidak bergantung pada distribusi data (Meidianingsih et al., n.d.). *Local Outlier Factor* memiliki kemampuan untuk mengidentifikasi anomali lokal maupun global yang tidak dapat diidentifikasi dengan pendekatan yang sudah ada [1].

Dalam pemantauan gunung api, anomali dapat muncul sebagai perubahan tiba-tiba dalam data seismik, deformasi tanah, emisi gas, atau aktivitas lain yang menandakan potensi erupsi. Gunung api Marapi yang terletak di Sumatera Barat sering meletus sejak tahun 1807, dengan masa istirahat erupsi bervariasi antara kurang dari 1 tahun hingga 17 tahun, dan rata-rata 3,5 tahun. Erupsi adalah proses di mana material seperti lava, gas, dan abu dikeluarkan dari dalam bumi ke permukaan melalui gunung api (Nurfadilah et al., 2023).

Pada 3 Desember 2023, terjadi erupsi yang tidak didahului oleh peningkatan gempa vulkanik yang signifikan. Gempa vulkanik dalam (VA) hanya terekam 3 kali antara tanggal 16 November 2023-2 Desember 2023 [2]. Letusan terjadi ketika banyak pendaki sedang berada di lereng gunung. Akibatnya, dari 75 pendaki yang mendaki gunung tersebut pada tanggal 2 Desember, 24 diantaranya meninggal dunia. Korban meninggal dunia ditemukan di dekat kawah gunung berapi.

Deteksi dini anomali ini penting untuk memberikan peringatan dini kepada pihak berwenang dan masyarakat sekitar tentang potensi bahaya erupsi. Pada penelitian ini akan diterapkan algoritma *Local Outlier Factor* untuk mendeteksi anomali pada data aktivitas kegempaan Gunung Marapi periode Oktober 2023 hingga Januari 2024. Berdasarkan uraian yang telah dikemukakan, maka rumusan masalah pada penelitian ini adalah sebagai berikut: (1) Berapa banyak data yang terdeteksi anomali dengan menggunakan algoritma *Local Outlier Factor*?. (2) Bagaimana karakteristik data yang terdeteksi anomali?. Adapun tujuan penelitian ini adalah untuk mengetahui jumlah dan karakteristik data yang terdeteksi anomali dari kegempaan yang terjadi pada Gunung Marapi. Melalui penelitian ini diharapkan dapat meningkatkan upaya monitoring dan mitigasi bencana Gunung Marapi, sehingga dapat meminimalisir dampak dan korban jiwa akibat erupsi.

## B. Metodologi Penelitian

Data yang dijadikan bahan penelitian ini merupakan data sekunder yang diperoleh dari *website magma.esdm.go.id*. Dataset ini mencakup informasi terkait aktivitas kegempaan Gunung Marapi pada periode Oktober 2023 hingga Januari 2024. Perangkat lunak yang digunakan dalam

penelitian ini adalah *Python 3.12 (64 bit) version* dengan menggunakan modul *StandardScaler* untuk untuk standarisasi data, *calculate\_kmo* untuk uji kecukupan sampel, *variance\_inflation\_factor* untuk uji non multikolinearitas, *KMeans* untuk analisis *K-Means clustering*, *KElbowVisualizer* untuk visualisasi grafik *Elbow*, dan *LocalOutlierFactor* untuk deteksi anomali. Unit penelitian yang digunakan adalah tanggal kejadian kegempaan yang terjadi di Gunung Marapi dan terdiri dari 123 pengamatan serta 9 variabel penelitian. Berikut adalah variabel penelitian yang digunakan.

**Tabel 1.** Variabel Penelitian

Variabel	Deskripsi Variabel	Skala
$X_1$	Frekuensi gempa vulkanik dalam	Rasio
$X_2$	Nilai maksimum S-P gempa vulkanik dalam (detik)	Rasio
$X_3$	Nilai maksimum amplitudo gempa vulkanik dalam (mm)	Rasio
$X_4$	Durasi maksimum gempa vulkanik dalam (detik)	Rasio
$X_5$	Frekuensi gempa vulkanik dangkal	Rasio
$X_6$	Nilai maksimum S-P gempa vulkanik dangkal (detik)	Rasio
$X_7$	Nilai maksimum amplitudo gempa vulkanik dangkal (mm)	Rasio
$X_8$	Durasi maksimum gempa vulkanik dangkal (detik)	Rasio
$X_9$	Frekuensi letusan/erupsi	Rasio

Pada penelitian ini terdapat beberapa tahapan analisis data yang terdiri dari analisis deskriptif, standarisasi data, penerapan *Elbow* pada *K-Means clustering*, dan deteksi anomali menggunakan *Local Outlier Factor (LOF)*. Analisis deskriptif dilakukan untuk menggambarkan karakteristik data aktivitas kegempaan Gunung Marapi. Standarisasi data dengan *z-score* dilakukan sebelum uji asumsi kluster untuk memastikan bahwa semua variabel berada pada skala yang sama. Kemudian diterapkan metode *Elbow* pada *K-Means* untuk mendapatkan nilai *K* yang optimal. Setelah menentukan nilai *K* optimal, algoritma *Local Outlier Factor (LOF)* digunakan untuk mendeteksi anomali dalam data.

### Analisis Kluster

Analisis kluster merupakan suatu teknik multivariat untuk mengelompokkan individu atau objek ke dalam kluster-kluster tersebut dengan sedemikian rupa. Sehingga objek-objek yang ada di dalam kluster yang sama atau mirip satu sama lain itu yang nantinya dibandingkan dengan objek-objek atau kluster lainnya [3]. Pada analisis kluster terdapat dua asumsi yang harus dipenuhi, yaitu sampel representatif atau kecukupan sampel dan tidak ada multikolinearitas [4].

### K-Means Clustering

*K-Means* adalah upaya untuk mengelompokkan atau membagi data yang ada secara non-hierarki di mana data yang memiliki karakteristik serupa ditempatkan dalam satu kluster yang sama, sementara data dengan karakteristik yang berbeda ditempatkan dalam kelompok yang berbeda [5]. Berikut merupakan algoritma dari *K-Means*:

1. Menentukan jumlah kluster ( $k$ ) yang ingin dibentuk.
2. Menetapkan  $k$  *centroid* (titik pusat kluster) awal ditentukan secara acak.

3. Menghitung jarak antara setiap data dengan pusat kluster (*centroid*) dari masing-masing kluster menggunakan jarak *euclidean*.
4. Kelompokkan data ke dalam kluster yang memiliki jarak paling pendek atau jarak paling dekat antara data dengan *centroid*nya.
5. Menghitung ulang pusat kluster (*centroid*) berdasarkan rata-rata data pada setiap kluster. Nilai *centroid* baru dapat dihitung menggunakan rumus berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} d_i \quad \dots (2.1)$$

Keterangan:  $C_k$ : pusat kluster baru dari kluster ke- $k$ ;  $n_k$ : jumlah data yang termasuk dalam kluster ke- $k$ ;  $d_i$ : data ke- $i$  yang termasuk dalam kluster ke- $k$

6. Lakukan perulangan dari langkah 3 hingga 5 sampai tidak ada lagi perpindahan data ke kluster lain. Setelah kondisi tersebut terpenuhi, maka parameter untuk menentukan klasifikasi data akan menggunakan nilai rata-rata pusat kluster pada iterasi terakhir.

### Metode Elbow

Pendekatan *Elbow* dapat digunakan untuk menemukan jumlah kelompok yang optimal dengan membandingkan persentase kelompok yang akan membentuk sudut pada titik tertentu dengan menghitung SSE (*Sum of Square Error*) untuk setiap nilai  $K$ . Nilai SSE akan menurun seiring bertambahnya jumlah kelompok  $K$  [6]. Berikut merupakan rumus SSE.

$$SSE = \sum_{j=1}^K \sum_{x_i \in C_j} |x_i - c_j|^2 \quad \dots (2.2)$$

SSE mengukur jumlah dari kuadrat jarak antara setiap titik data ( $x_i$ ) dalam sebuah kluster ( $C_j$ ) dengan pusat kluster ( $c_j$ ). Pusat kluster ( $c_j$ ) diperoleh dengan menghitung rata-rata dari semua titik data yang termasuk dalam kluster ( $C_j$ ) yang dapat dihitung menggunakan rumus pada Persamaan (2.1). Kemudian grafik garis SSE digambarkan untuk setiap nilai  $k$ . Jika grafik garis terlihat seperti lengan maka “siku” pada lengan tersebut memberikan nilai  $k$  yang terbaik [7].

### Algoritma *Local Outlier Factor* (LOF)

*Local Outlier Factor* adalah algoritma pembelajaran mesin yang digunakan untuk mendeteksi *outlier* lokal dalam suatu dataset yang lebih berfokus pada mendeteksi titik data yang berbeda dari tetangga lokalnya daripada mendeteksi *outlier* global yang terlihat berbeda dari keseluruhan dataset [1]. LOF bekerja dengan membandingkan kerapatan suatu titik data dengan kerapatan titik-titik di sekitarnya, di mana titik data dengan kerapatan jauh lebih rendah dari tetangganya atau data yang memiliki skor LOF yang tinggi akan dianggap sebagai anomali. Adapun definisi yang perlu diketahui dalam algoritma *Local Outlier Factor* yaitu [1]:

**Definisi 1.**  $k$ -distance( $p$ ), jarak antara suatu titik dan merupakan tetangga terdekat ke- $k$ .

**Definisi 2.**  $k$ -distance neighborhood( $p$ ), himpunan tetangga dari suatu objek  $p$  yang didefinisikan berdasarkan jarak  $k$ -distance( $p$ ). Setiap titik data  $q$  yang jaraknya ke titik data  $p$  tidak lebih besar dari  $k$ -distance( $p$ ).

$$N_{k\text{-distance}(p)} = \{q \in D\{p\} | d(p, q) \leq k - \text{distance}(p)\}$$

**Definisi 3.** *Reachability Distance*( $p$ ), jarak maksimum antara objek dengan tetangga terdekat ke- $k$ -nya dan jarak antara objek  $o$  dengan objek  $p$ .

$$\text{reach - dist}_k(p, o) = \max \{k - \text{distance}(o), d(p, o)\} \quad \dots (2.3)$$

**Definisi 4.** *Local Reachability Density*( $p$ ), kebalikan dari rata-rata *reachability distance* berdasarkan  $k$ -nearest neighbors dari  $p$ .

$$lrd_k(p) = \frac{1}{\left( \frac{\sum_{o \in N_k(p)} reach-dist_k(p,o)}{|N_k(p)|} \right)} \quad \dots (2.4)$$

Keterangan:  $lrd_k(p)$  : kerapatan jangkauan lokal dari objek  $p$  dengan parameter  $k$ ;  $\sum_{o \in N_k(p)} reach - dist_k(p, o)$  : jumlah dari jarak jangkauan ( $reach - dist_k$ ) antara objek  $p$  dan setiap tetangga  $o$  dari  $p$  dalam radius  $k$ ;  $|N_k(p)|$  : jumlah tetangga terdekat  $p$  dalam radius  $k$

**Definisi 5.** *Local Outlier Factor*( $p$ ), perbandingan rata-rata *local reachability density* dari  $p$  dengan *k-nearest neighbors* dari  $p$ .

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|} \quad \dots (2.5)$$

Keterangan:  $LOF_k(p)$  : *Local Outlier Factor* dari objek  $p$  dengan parameter  $k$ ;  $N_k(p)$  : himpunan tetangga terdekat  $p$  dalam radius  $k$ ;  $\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}$  : jumlah dari rasio kerapatan jangkauan lokal ( $lrd$ ) dari setiap tetangga  $o$  dari  $p$  dalam radius  $k$  dibagi dengan  $lrd$  dari  $p$ ;  $|N_k(p)|$  : jumlah tetangga terdekat  $p$  dalam radius  $k$

Berikut aturan ambang batas yang diberikan [8]:

1. Jika skor LOF  $\sim 1$ , itu berarti kerapatan serupa dengan tetangganya.
2. Jika skor LOF  $< 1$ , itu berarti kerapatan lebih tinggi daripada tetangganya, yang disebut *inlier*.
3. Jika skor LOF  $> 1$ , itu berarti kerapatan lebih rendah daripada tetangganya, yang disebut *outlier*.

Adapun tahapan analisis yang dilakukan pada penelitian ini sebagai berikut:

1. Mengumpulkan data aktivitas kegempaan Gunung Marapi dari *website magma.esdm.go.id* selama periode Oktober 2023 hingga Januari 2024.
2. Standarisasi data menggunakan *z-score* dengan rumus:

$$z_i = \frac{x_i - \bar{x}}{s} \quad \dots (2.6)$$

Keterangan:  $x_i$  : nilai atau objek ke- $i$ ;  $\bar{x}$  : rata-rata sampel;  $s$  : standar deviasi dari seluruh data

3. Uji asumsi kecukupan sampel dengan melihat nilai *Kaiser-Meyer-Olkin* (KMO) dan asumsi non multikolinearitas dengan melihat nilai *Variance Inflation Factor* (VIF).
4. Menentukan nilai  $K$  optimal menggunakan metode *Elbow* yang diterapkan pada *K-Means*.
5. Menghitung jarak dengan tetangga terdekat menggunakan jarak *euclidean* dengan rumus:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_i^{(k)} - c_j^{(k)})^2} \quad \dots (2.7)$$

Keterangan:

$d(x_i, c_j)$  : jarak antara titik data  $x_i$  dan pusat kluster  $c_j$ ;  $n$  : jumlah dimensi dari titik data  $x_i$  dan pusat kluster  $c_j$ ;  $x_i^{(k)}$  : elemen ke- $k$  dari titik data  $x_i$ ;  $c_j^{(k)}$  : elemen ke- $k$  dari pusat kluster  $c_j$

6. Menghitung kerapatan jangkauan lokal (*local reachability density*) dengan rumus Persamaan (2.4).
7. Menghitung skor *Local Outlier Factor* dengan rumus Persamaan (2.5).
8. Mengurutkan skor *Local Outlier Factor* dari yang terbesar hingga terkecil di mana data yang memiliki skor LOF lebih besar dari ambang batas yang telah ditentukan maka data

- tersebut terdeteksi sebagai anomali.
- 9. Mendapatkan data yang terdeteksi anomali.
- 10. Mendeskripsikan karakteristik data yang terdeteksi anomali.
- 11. Menarik kesimpulan.
- 12. Metodologi Penelitian

**C. Hasil Penelitian dan Pembahasan**

**Karakteristik Data Aktivitas Kegempaan Gunung Marapi**

Berikut merupakan karakteristik dari kegempaan yang terjadi pada Gunung Marapi.

**Tabel 2.** Statistik Deskriptif Kegempaan Gunung Marapi

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
<b>Rata-rata</b>	0,561	0,059	1,797	4,870	0,333	0,699	0,420	1,211	1,130
<b>Standar Deviasi</b>	1,065	0,265	4,866	8,832	2,462	4,787	2,046	4,289	3,695
<b>Maksimum</b>	5	1,5	30,4	55	27	41	20	26	36

Gunung Marapi menunjukkan variasi aktivitas kegempaan selama periode Oktober 2023 hingga Januari 2024. Gempa vulkanik dalam terjadi dengan frekuensi rata-rata 0,561 kali per hari dengan variasi yang tidak tinggi, namun beberapa hari menunjukkan aktivitas yang lebih tinggi, dengan frekuensi maksimum mencapai 5 gempa dalam sehari. Gelombang S-P rata-rata membutuhkan 0,059 detik untuk mencapai stasiun pengamat, dengan amplitudo maksimum mencapai 30,4 mm, dan durasi maksimum 55 detik.

Gempa vulkanik dangkal terjadi dengan frekuensi rata-rata yang lebih rendah yaitu 0,333 kali per hari, dengan variasi yang besar tercatat hingga 27 gempa dalam satu hari. Rentang waktu S-P untuk gempa vulkanik dangkal bisa mencapai hingga 41 detik, dengan amplitudo maksimum 20 mm dan durasi maksimum 26 detik, menunjukkan bahwa aktivitas gempa vulkanik dangkal dapat meningkat secara signifikan pada hari-hari tertentu.

Letusan atau erupsi terjadi dengan frekuensi rata-rata 1,130 kali per hari, dengan frekuensi maksimum mencapai 36 erupsi dalam sehari. Variasi besar dalam aktivitas erupsi ini menunjukkan bahwa Gunung Marapi dapat memiliki hari-hari dengan aktivitas vulkanik yang sangat intens, berpotensi menimbulkan risiko bagi masyarakat sekitar.

Secara keseluruhan, gempa vulkanik dalam dan vulkanik dangkal di Gunung Marapi menunjukkan aktivitas vulkanik yang cenderung rendah dengan frekuensi yang cenderung stabil. Waktu tempuh gelombang S-P yang pendek menunjukkan bahwa gempa ini berasal dari sumber yang dekat dengan stasiun pengamat. Getaran yang dihasilkan umumnya lemah, namun beberapa gempa menghasilkan getaran yang cukup kuat dan berlangsung dengan durasi gempa yang tidak terlalu lama, yaitu kurang dari satu menit.

Aktivitas vulkanik di Gunung Marapi sangat bervariasi dari hari ke hari, dengan potensi peningkatan aktivitas yang signifikan pada waktu-waktu tertentu. Hal ini menekankan pentingnya pemantauan terus-menerus untuk mengantisipasi dan mengurangi risiko bencana vulkanik yang dapat terjadi.

**Uji Asumsi Analisis Klaster**

Dalam analisis klaster, penting untuk memastikan bahwa data memenuhi beberapa asumsi dasar sebelum melanjutkan ke langkah-langkah klasterisasi. Terdapat dua asumsi yang harus dipenuhi untuk menguji kelayakan data dalam analisis klaster, yaitu uji kecukupan sampel atau sampel mewakili populasi dan uji non multikolinearitas. Sebelum melakukan kedua uji asumsi ini, data harus distandarasi terlebih dahulu menggunakan *z-score*. Proses standarisasi ini bertujuan untuk mengubah skala variabel sehingga semua variabel memiliki skala yang seragam untuk menghindari bias dalam analisis klaster. Standarisasi data pada *Python* dapat dilakukan

menggunakan *StandardScaler* dari *library scikit-learn*. Hasil dari standarisasi data disajikan pada Tabel 3.

**Tabel 3.** Hasil Standarisasi Data

No.	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
1	-0,529	-0,224	-0,371	-0,554	-0,136	-0,147	-0,206	-0,284	-0,307
2	-0,529	-0,224	-0,371	-0,554	-0,136	-0,147	-0,206	-0,284	-0,307
3	-0,529	-0,224	-0,371	-0,554	-0,136	-0,147	-0,206	-0,284	-0,307
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
121	-0,529	-0,224	-0,371	-0,554	0,680	6,922	3,033	2,057	-0,307
122	-0,529	-0,224	-0,371	-0,554	0,272	8,453	1,168	1,589	-0,307
123	0,414	-0,224	-0,061	0,981	0,272	0,315	0,334	2,760	-0,307

#### Uji Sampel Representatif

Setelah data di standarisasi maka dilakukan pengujian untuk mengukur kecukupan sampel dan untuk mengetahui apakah data yang ada cocok untuk analisis kluster dengan melihat nilai *Kaiser-Meyer-Olkin* (KMO). Sampel dikatakan telah mewakili populasi apabila nilai KMO > 0,5. Pengujian KMO pada *Python* dapat dilakukan menggunakan *library factor\_analyzer* dengan modul *calculate\_kmo*. Berikut merupakan tabel hasil pengujian kecukupan sampel dengan *Kaiser-Meyer-Olkin* (KMO).

**Tabel 4.** Statistik Uji KMO

<i>Kaiser-Meyer-Olkin</i>	0,616
---------------------------	-------

Berdasarkan pengujian yang telah dilakukan di dapat hasil bahwa nilai KMO (0,616) > 0,5. Hal tersebut menunjukkan bahwa asumsi kecukupan sampel atau sampel mewakili populasi telah terpenuhi karena nilai KMO > 0,5.

#### Uji Non Multikolinearitas

Setelah asumsi sampel mewakili populasi terpenuhi maka langkah selanjutnya adalah melakukan uji asumsi non multikolinieritas. Uji asumsi non multikolinearitas dilakukan untuk mengetahui hubungan antar variabel bebas yang digunakan. Uji asumsi non multikolinearitas dapat dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF). Asumsi non multikolinearitas terpenuhi jika nilai VIF < 10. Pengujian VIF pada *Python* dapat dilakukan menggunakan *library statsmodels* dengan modul *variance\_inflation\_factor*. Berikut merupakan tabel hasil pengujian non multikolinieritas dengan *Variance Inflation Factor* (VIF).

**Tabel 5.** Statistik Uji VIF

No.	Variabel	VIF	No.	Variabel	VIF
1	$X_1$	2,882	6	$X_6$	1,179
2	$X_2$	1,065	7	$X_7$	1,440

No.	Variabel	VIF	No.	Variabel	VIF
3	$X_3$	1,351	8	$X_8$	2,315
4	$X_4$	2,852	9	$X_9$	1,016
5	$X_5$	1,836			

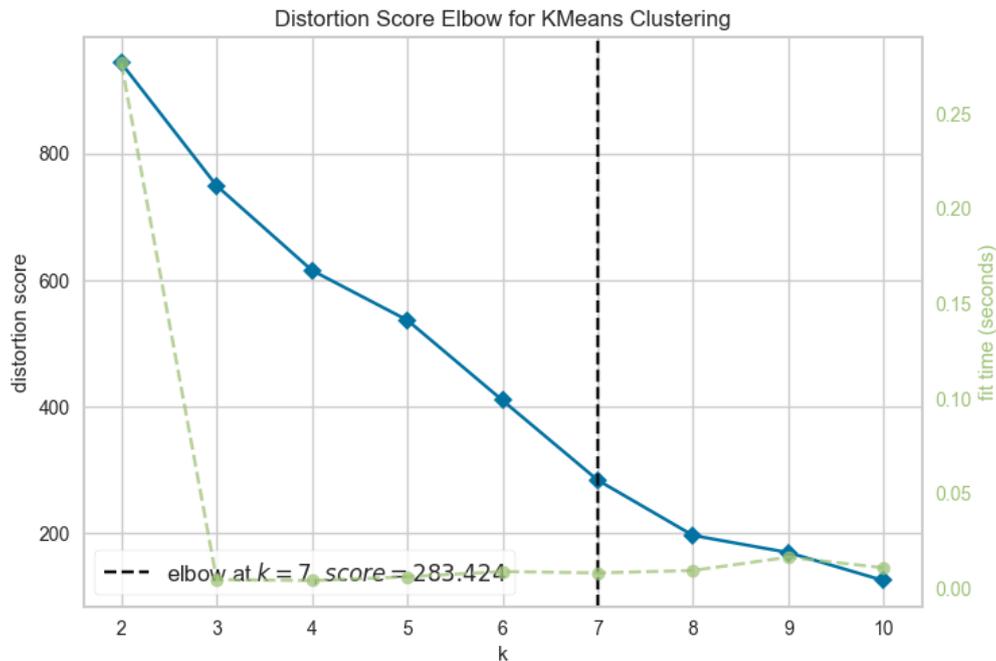
Berdasarkan pengujian yang telah dilakukan di dapat hasil bahwa nilai VIF untuk seluruh variabel bebas kurang dari 10 ( $VIF < 10$ ). Hal tersebut menunjukkan bahwa tidak terdapat multikolinieritas antar variabel bebas atau asumsi non multikolinieritas telah terpenuhi.

### Deteksi Anomali dengan Local Outlier Factor

Selanjutnya, akan dilakukan deteksi anomali menggunakan algoritma *Local Outlier Factor* (LOF) dengan bantuan bahasa pemrograman *Python* dan diterapkan pada data aktivitas kegemampuan Gunung Marapi periode Oktober 2023 hingga Januari 2024. Adapun hal yang harus diperhatikan adalah menentukan nilai  $K$  optimal dan nilai ambang batas di mana penentuan ambang batas dilakukan secara subjektif oleh peneliti. Langkah pertama adalah menentukan nilai  $K$  optimal menggunakan metode *Elbow*. Nilai  $K$  optimal didapatkan berdasarkan evaluasi nilai *Sum of Squared Errors* (SSE) dari hasil *clustering* menggunakan algoritma *K-Means* di mana penentuan nilai  $K$  dilakukan dengan percobaan 2 hingga 10 klaster. Algoritma *K-Means* diimplementasikan menggunakan *library scikit-learn* dengan menggunakan modul *KMeans*, sedangkan visualisasi grafik *Elbow* atau nilai  $K$  optimal diimplementasikan menggunakan *library yellowbrick* dengan modul *KElbowVisualizer*. Berikut merupakan perhitungan dan grafik dari SSE.

**Tabel 6.** Hasil Perhitungan SSE

Banyaknya Cluster (k)	<i>Sum of Squared Error</i> (SSE)	Banyaknya Cluster (k)	<i>Sum of Squared Error</i> (SSE)
2	942,973	7	283,424
3	749,168	8	196,061
4	615,303	9	169,052
5	536,667	10	124,662
6	409,717		



**Gambar 1.** Grafik SSE

Berdasarkan grafik di atas maka dapat disimpulkan bahwa nilai  $K$  optimal adalah 7 atau  $K = 7$ .

Setelah nilai  $K$  optimal ditentukan, dilakukan proses deteksi anomali menggunakan algoritma *Local Outlier Factor* (LOF). Algoritma ini diimplementasikan menggunakan *library scikit-learn* dengan menggunakan modul *LocalOutlierFactor*. Algoritma LOF bekerja dengan menghitung skor LOF untuk setiap titik data. Skor LOF ini mewakili tingkat keanehan suatu titik data dibandingkan dengan titik data lain di sekitarnya. Titik data dengan skor LOF tinggi dianggap sebagai anomali, sedangkan titik data dengan skor LOF rendah dianggap normal. Untuk menentukan apakah suatu titik data dikategorikan sebagai anomali, diperlukan ambang batas skor LOF. Ambang batas ini dapat ditentukan secara subjektif oleh peneliti sesuai dengan kebutuhan dan konteks penelitian. Pada penelitian ini, digunakan ambang batas 1. Jika skor LOF lebih dari 1 maka titik data tersebut dikategorikan sebagai anomali. Hasil deteksi anomali disajikan pada Tabel 6 di bawah ini.

**Tabel 6.** Hasil Deteksi Anomali

No.	Tanggal	Kategori	No.	Tanggal	Kategori
1	10/1/2023	Bukan Anomali	⋮	⋮	⋮
2	10/2/2023	Bukan Anomali	121	1/29/2024	Anomali
3	10/3/2023	Bukan Anomali	122	1/30/2024	Anomali
⋮	⋮	⋮	123	1/31/2024	Anomali

Tabel 6 menunjukkan terdapat 43 kejadian yang dideteksi sebagai anomali menggunakan algoritma *Local Outlier Factor*. Selanjutnya, perlu dilihat karakteristik dari data yang terdeteksi sebagai anomali.

**Karakteristik Data yang Terdeteksi Anomali**

Berikut merupakan karakteristik data yang terdeteksi anomali dari analisis yang telah dilakukan terhadap 123 data aktivitas kegempaan Gunung Marapi.

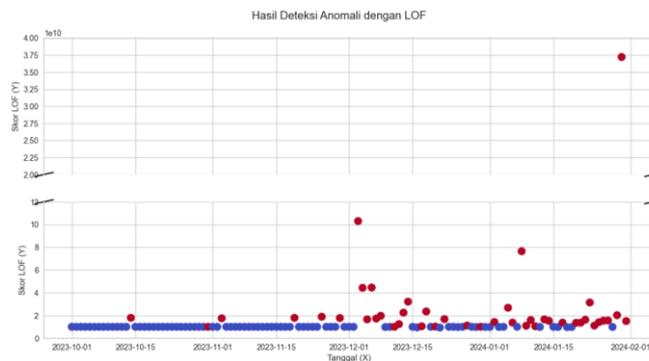
**Tabel 7.** Data yang Terdeteksi Anomali

No.	Tanggal	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	LOF
1	10/14/2023	2	1,2	4,0	12,0	0	0	0	0	0	1,791
2	10/31/2023	1	0	1,1	11,0	0	0	0	0	0	1,000
3	11/3/2023	1	1,0	2,6	12,0	0	0	0	0	0	1,751
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	1/29/2024	0	0	0	0	2	33,7	6,6	10	0	2,018
42	1/30/2024	0	0	0	0	1	41,0	2,8	8	0	3,722 × 10 <sup>10</sup>
43	1/31/2024	1	0	1,5	13,5	1	2,2	1,1	13	0	1,505

Berdasarkan data keseluruhan, terdapat 43 atau 34,96% kejadian yang terdeteksi sebagai anomali. Berdasarkan statistik deskriptif pada Tabel 2, rata-rata frekuensi gempa vulkanik dalam adalah 0,561 per hari, sedangkan dalam data anomali, frekuensinya mencapai nilai maksimum hingga 5 kejadian gempa dalam satu hari. Rata-rata waktu S-P untuk gempa vulkanik dalam adalah 0,059 detik, tetapi dalam data anomali terdapat nilai hingga 1,5 detik. Amplitudo maksimum gempa vulkanik dalam pada data anomali mencapai 30,4 mm, sesuai dengan maksimum keseluruhan, menunjukkan bahwa gempa yang kuat sering kali terdeteksi sebagai anomali. Durasi maksimum gempa vulkanik dalam pada data anomali juga mencapai 55 detik, yang merupakan durasi maksimum dari keseluruhan data.

Sedangkan gempa vulkanik dangkal dalam data anomali memiliki frekuensi yang tinggi mencapai 27 kejadian gempa dalam satu hari, dibandingkan dengan rata-rata keseluruhan yang hanya 0,333 per hari. Waktu S-P untuk gempa vulkanik dangkal dalam data anomali mencapai hingga 41 detik, jauh lebih tinggi dibandingkan rata-rata keseluruhan yang hanya 0,699 detik. Selain itu, amplitudo maksimum untuk gempa vulkanik dangkal dalam data anomali juga mencapai nilai maksimum hingga 20 mm, dibandingkan dengan rata-rata keseluruhan yang hanya 0,420 mm. Durasi maksimum gempa vulkanik dangkal pada data anomali mencapai 26 detik, yang juga merupakan durasi maksimum dari keseluruhan data.

Letusan atau erupsi dalam data anomali terjadi dengan frekuensi yang lebih tinggi, dengan nilai maksimum mencapai 36 kejadian dalam satu hari, dibandingkan dengan rata-rata keseluruhan 1,130 per hari. Hasil deteksi anomali juga ditampilkan secara visual yang tersaji pada Gambar 2.



**Gambar 2.** Visualisasi skor LOF

Berdasarkan visualisasi di atas, titik data yang terdeteksi sebagai anomali ditandai dengan warna merah. Beberapa anomali memiliki skor LOF yang sangat tinggi, yaitu 4,5 hingga 10,3 yang terlihat pada bagian bawah grafik (subplot kedua) dan mencapai  $3,7 \times 10^{10}$  yang terlihat pada bagian atas grafik (subplot pertama). Hal itu menunjukkan adanya perbedaan kerapatan yang signifikan dibandingkan dengan titik data lainnya. Jika dilihat dari tanggal kejadian, algoritma *Local Outlier Factor* (LOF) berhasil mendeteksi adanya anomali pada tanggal 3, 19, 25, dan 29 November 2023, dengan deteksi terakhir sekitar empat hari sebelum erupsi yang terjadi pada tanggal 3 Desember 2023. Data yang terdeteksi anomali juga umumnya memiliki tanggal kejadian yang berdekatan bahkan berurutan, seperti yang terlihat dari beberapa kejadian anomali yang berurutan pada tanggal 3 hingga 8 Desember 2023, 11 hingga 14 Desember 2023, dan 20 hingga 27 Januari 2024. Ini menunjukkan pola yang mengindikasikan adanya peningkatan aktivitas vulkanik yang lebih signifikan dalam periode waktu tersebut. Deteksi ini menunjukkan bahwa metode LOF dapat memberikan peringatan dini mengenai peningkatan aktivitas vulkanik, sehingga memungkinkan tindakan pencegahan yang lebih baik dan mengurangi risiko bagi masyarakat di sekitar Gunung Marapi.

#### D. Kesimpulan

Berdasarkan analisis yang telah dilakukan menggunakan algoritma *Local Outlier Factor* (LOF), dapat disimpulkan bahwa:

1. Sebanyak 43 kejadian gempa terdeteksi anomali atau 34,96% dari total data aktivitas kegempaan Gunung Marapi.
2. Data anomali umumnya memiliki karakteristik berupa nilai frekuensi, S-P, amplitudo, dan durasi lebih tinggi daripada rata-rata keseluruhan, meskipun ada beberapa data anomali yang berada di bawah rata-rata, namun ada juga yang mendekati nilai maksimum atau memiliki nilai yang tinggi pada parameter lain. LOF berhasil mendeteksi anomali pada tanggal 29 November 2023, sekitar empat hari sebelum erupsi yang terjadi pada tanggal 3 Desember 2023. Hal itu menunjukkan kemampuan LOF dalam memberikan peringatan dini dan memungkinkan tindakan pencegahan yang lebih baik untuk mengurangi risiko bagi masyarakat di sekitar Gunung Marapi. Data yang terdeteksi anomali umumnya juga memiliki tanggal atau waktu kejadian yang berdekatan atau berurutan.

#### Acknowledge

Segala puji dan syukur kami panjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya yang telah memberikan kekuatan dan kemudahan dalam menyelesaikan penelitian ini. Terima kasih kepada Bapak Dr. Nusar Hajarisman, M.Si. sebagai dosen pembimbing yang telah memberikan bimbingan, saran, dan koreksi yang sangat berarti selama penelitian ini. Ucapan terima kasih juga kami sampaikan kepada para dosen Program Studi Statistika di Universitas Islam Bandung atas ilmu dan dukungan yang telah diberikan, serta kepada semua pihak yang telah membantu dan memberikan dukungan baik dalam bentuk bantuan teknis maupun motivasi selama proses penelitian ini. Semoga hasil penelitian ini bermanfaat dan dapat memberikan kontribusi yang berarti bagi perkembangan ilmu pengetahuan.

#### Daftar Pustaka

- [1] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 93–104. <https://doi.org/10.1145/342009.335388>
- [2] PVMBG. (2024). *Penyampaian kenaikan tingkat aktivitas G. Marapi, Sumatera Barat dari Level II (Waspada) menjadi Level III (Siaga)*. [https://vsi.esdm.go.id/files/756/71%20Peningkatan%20Status%20G.%20Marapi%20dari%20Waspada%20menjadi%20Siaga\\_9%20Januari%202024%20\(1\)%20.pdf](https://vsi.esdm.go.id/files/756/71%20Peningkatan%20Status%20G.%20Marapi%20dari%20Waspada%20menjadi%20Siaga_9%20Januari%202024%20(1)%20.pdf)
- [3] Hajarisman, N. (2019). *Statistika Multivariat: Analisis Kluster*. Bandung: NH Press.
- [4] Santoso, F. (2023). *ANALISIS KLASSTER INFRASTRUKTUR SEKTOR KESEHATAN DI WILAYAH KABUPATEN WONOGIRI TAHUN 2021*

- MENGGUNAKAN METODE ALGORITMA K-MEANS. *Inisiasi*, 7–16. <https://doi.org/10.59344/inisiasi.v12i1.118>
- [5] Murti, M. (2017). Penerapan metode K-means clustering untuk mengelompokan potensi produksi buah–buahan di provinsi daerah istimewa yogyakarta. *Universitas Sanata Dharma*.
- [6] Merliana, N. P. E., Ernawati, & Santoso, A. J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering. *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI\_U)*.
- [7] Duong, M. Q., Lam, L. H., Tu, B. T. M., Huy, G. Q., & Hieu, N. H. (2019). A Combination of K-Mean Clustering and Elbow Technique in Mitigating Losses of Distribution Network. *GMSARN International*, 13, 153–158.
- [8] Zadafiya, N., Karasariya, J., Kanani, P., & Nayak, A. (2022). Detecting Credit Card Frauds Using Isolation Forest And Local Outlier Factor - Analytical Insights. *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 1588–1594. <https://doi.org/10.1109/ICSSIT53264.2022.9716541>
- [9] Agustin Nuriani Sirodj, D., Made Sumertajaya, I., & Kurnia, A. (2023). Analisis Clustering Time Series untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Manusia Jenis Kelamin Perempuan (Vol. 23, Issue 1). <https://www.bps.go.id/indicator/40/462/1/indeks-pembangunan-manusia-ipm-menurut->
- [10] Meidianingsih, Q., Wardani, D. E., Salsabila, E., Nafisah, L., & Mutia, A. N. (n.d.). Perbandingan Performa Metode Berbasis Support Vector Machine untuk Penanganan Klasifikasi Multi Kelas Tidak Seimbang (Vol. 23, Issue 1).
- [11] Nurfadilah, K., Notodiputro, K. A., Sartono, B., & Nas, A. (2023). Premarital Sex Behavior Model with Lasso Generalized Linear Mixed Model and Group Lasso Generalized Linear Mixed Model (Vol. 23, Issue 1).