PENERAPAN ALGORITMA CLUSTERING BIRCH DAN DBSCAN DALAM ANALISIS GEMPA BUMI DI INDONESIA

Muhammad Yusuf[1], Asyrofi Anam[2], Muhammad Rifki Raihan[3]

[1,2,3]Program Studi Teknik Informatika, Universitas Muhammadiyah Sorong

Jln. Pendidikan No.27, Kelurahan Klabulu, Malaimsimsa, Kota Sorong, Papua Barat Daya 98416

[1] yusuf@um-sorong.ac.id, [2] asyrofianam0@gmail.com, [3] r2928971@gmail.com

Abstract

Earthquakes are natural phenomena that occur due to the shifting of the earth's plates, fluid activity, volcanic activity, and active faults, and are destructive because they often cause major losses. The impacts include infrastructure damage, economic losses, and casualties. Therefore, identifying earthquake-prone areas is a crucial step in increasing awareness. This study aims to explore and compare the performance of the BIRCH and DBSCAN clustering algorithms in analyzing earthquake data in Indonesia until 2025. BIRCH was chosen because it is efficient for large data, while DBSCAN excels in detecting noise and irregular patterns. Evaluation was carried out using the Silhouette coefficient, Davies-Bouldin index, and Calinski-Harabasz index metrics. Visualization was performed through PCA and spatial mapping. The results show that BIRCH with two clusters provides the best results with a Silhouette coefficient of 0.8910, a Davies-Bouldin index of 0.3580, and a Calinski-Harabasz index of 1009.83. Meanwhile, DBSCAN produced a Silhouette coefficient of 0.8726, a Davies-Bouldin index of 0.5017, and a Calinski-Harabasz index of 554.37. BIRCH proved to be more stable and representative in depicting the distribution of earthquake risk. The advantages of this research open up opportunities for further development, such as adding analytical attributes and evaluation approaches to improve the accuracy of classifying earthquake-prone areas in Indonesia.

Keywords : *earthquake, data mining, clustering, BIRCH, DBSCAN*

Abstrak

Gempa bumi merupakan fenomena alam yang terjadi akibat pergeseran lempeng bumi, aktivitas fluida, vulkanik, maupun sesar aktif, dan bersifat destruktif karena sering menimbulkan kerugian besar. Dampaknya mencakup kerusakan infrastruktur, kerugian ekonomi, dan korban jiwa. Oleh karena itu, identifikasi kawasan rawan gempa menjadi langkah penting dalam meningkatkan kewaspadaan. Penelitian ini bertujuan untuk mengeksplorasi dan membandingkan performa algoritma *clustering* BIRCH dan DBSCAN dalam menganalisis data gempa bumi di Indonesia hingga tahun 2025. BIRCH dipilih karena efisien untuk data besar, sedangkan DBSCAN unggul dalam mendeteksi *noise* dan pola yang tidak beraturan. Hasil evaluasi menunjukkan bahwa BIRCH dengan dua kelompok mampu memberikan pemetaan yang lebih jelas dan stabil dibandingkan DBSCAN. Pada BIRCH, nilai pemisahan antar kelompok lebih tinggi (Silhouette 0,8910; Calinski-Harabasz 1009,83) dan tingkat kesalahannya lebih rendah (Davies-Bouldin 0,3580). Sebaliknya, DBSCAN menghasilkan pemisahan yang kurang optimal dengan skor yang lebih rendah (Silhouette 0,8726; Calinski-Harabasz 554,37; Davies-Bouldin 0,5017). BIRCH terbukti lebih stabil dan representatif dalam menggambarkan distribusi risiko gempa bumi. Keunggulan dari penelitian ini membuka peluang untuk dilakukan pengembangan lanjutan, seperti penambahan atribut analisis dan pendekatan evaluasi guna meningkatkan akurasi dalam pengelompokan wilayah rawan gempa di Indonesia.

Kata kunci : *gempa bumi, data mining, pengelompokan data, BIRCH, DBSCAN*

1. PENDAHULUAN

Gempa bumi merupakan suatu fenomena alam yang salah satunya terjadi akibat pergeseran lempeng pada permukaan bumi. Peristiwa ini bersifat *destruktif* karena hampir setiap kejadiannya menimbulkan kerugian baik secara *materiil* maupun *imateriil*. Selain pergeseran lempeng, gempa juga dapat dipicu oleh aktivitas fluida, vulkanik, maupun sesar aktif, dengan mekanisme sumber yang berbeda dibandingkan gempa besar akibat *subduksi megathrust* atau sesar utama [1]. Dampak dari gempa bumi dapat mencakup kerusakan infrastruktur, kerugian ekonomi, hingga menimbulkan korban jiwa. Oleh karena itu, identifikasi dan analisis terhadap kawasan rawan gempa menjadi langkah penting dalam mendukung upaya pencegahan dan peningkatan kewaspadaan masyarakat terhadap risiko bencana [2].

Seiring berkembangnya teknologi dan ketersediaan data gempa bumi yang semakin luas, pendekatan manual dalam menganalisis data gempa mulai tergantikan oleh metode otomatis berbasis *data mining*. Salah satu pendekatan yang efektif adalah *clustering*, yakni teknik yang mampu mengelompokkan data berdasarkan kesamaan karakteristik tanpa memerlukan label [3]. Metode ini dinilai lebih adaptif dalam menggambarkan pola tersembunyi dan potensi kerawanan wilayah berdasarkan data historis gempa bumi.

Namun, penerapan algoritma *clustering* dalam analisis gempa bumi bukan hal yang sederhana. Beberapa tantangan seperti ketidakseimbangan data antar wilayah, *noise* pada data lokasi, hingga variasi magnitudo dan kedalaman gempa sering menjadi hambatan dalam menghasilkan *cluster* yang representatif. Selain itu, masing-masing algoritma *clustering* memiliki cara kerja dan performa yang berbeda terhadap bentuk dan distribusi data. Oleh karena itu, penting untuk mengevaluasi dan membandingkan antar algoritma untuk memperoleh hasil klasifikasi wilayah rawan gempa yang lebih akurat dan informatif.

Beberapa penelitian sebelumnya telah mengevaluasi efektivitas berbagai algoritma *clustering* dalam analisis data gempa bumi di Indonesia. Salah satu studi membandingkan algoritma K-Medoids dan K-Means menggunakan perangkat lunak *R*, dan menghasilkan nilai *silhouette coefficient* sebesar 0,546 untuk K-Medoids dan 0,516 untuk K-Means, menunjukkan bahwa K-Medoids memiliki kemampuan pemisahan *cluster* yang sedikit lebih baik [4]. Penelitian lain mengelompokkan data gempa menggunakan algoritma K-Means dan DBSCAN. Dalam penelitian tersebut, DBSCAN membentuk 14 *cluster* dan mengidentifikasi 102 data sebagai *noise* menggunakan nilai *MinPts* sebesar 12 dan *Eps* sebesar 0,13. Meskipun nilai *silhouette coefficient* DBSCAN lebih tinggi yaitu 0,73 dibandingkan K-Means yang memperoleh 0,70, K-Means dinilai menghasilkan *cluster* yang lebih rapi dengan tingkat kesalahan yang lebih rendah [5].

Selain itu, pendekatan lain dengan algoritma K-Medoids menunjukkan hasil yang cukup baik dengan membentuk dua *cluster* utama dan nilai *silhouette coefficient* sebesar 0,68016 [6]. Penelitian lainnya yang berfokus pada wilayah Pulau Jawa membandingkan performa K-Means dan DBSCAN dalam bentuk visualisasi spasial melalui peta, dan menemukan bahwa K-Means memberikan hasil yang lebih baik dengan indeks *silhouette* 0,54 dibandingkan DBSCAN yang hanya mencapai 0,17 [7].

Walaupun hasil yang diperoleh dari masing-masing penelitian menunjukkan bahwa algoritma seperti K-Means dan K-Medoids memiliki kinerja yang cukup baik dalam mengelompokkan data gempa bumi, namun efektivitas algoritma sangat bergantung pada karakteristik data yang digunakan. Beberapa algoritma menunjukkan performa tinggi dalam hal nilai *silhouette coefficient*, tetapi kurang optimal dalam menangani *noise* atau data *outlier*. Sementara itu, algoritma seperti DBSCAN meskipun unggul dalam mengidentifikasi *noise*, cenderung menghasilkan *cluster* yang tidak seragam jika parameter tidak ditentukan secara tepat. Oleh karena itu, masih diperlukan analisis lebih lanjut dengan membandingkan algoritma lain yang mungkin lebih fleksibel dalam menangani data gempa yang kompleks dan tersebar secara spasial.

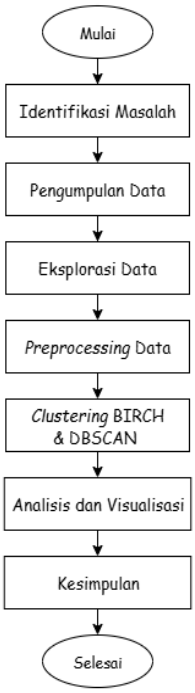
Sejumlah penelitian terdahulu mengenai analisis data gempa bumi di Indonesia umumnya menggunakan algoritma clustering konvensional seperti K-Means dan K-Medoids. Meskipun mampu membentuk kelompok data, algoritma tersebut masih memiliki keterbatasan dalam menghadapi data yang bising maupun pola distribusi yang tidak beraturan. Beberapa studi juga melibatkan DBSCAN dan menunjukkan keunggulannya dalam mendeteksi data outlier, namun hasil yang diperoleh sangat bergantung pada pemilihan parameter sehingga kurang stabil. Berangkat dari keterbatasan tersebut, penelitian ini hadir sebagai upaya untuk melengkapi studi sebelumnya dengan mengeksplorasi dan membandingkan performa

algoritma BIRCH dan DBSCAN dalam menganalisis data gempa bumi di Indonesia menggunakan data terbaru hingga tahun 2025. Kedua algoritma ini dipilih karena menawarkan pendekatan yang saling melengkapi: DBSCAN unggul dalam mengenali pola cluster yang tidak beraturan serta mengidentifikasi noise, sedangkan BIRCH lebih efisien dalam mengolah dataset besar melalui struktur pohon hierarki. Evaluasi dilakukan dengan tiga metrik utama, yaitu Silhouette coefficient, Davies-Bouldin index, dan Calinski-Harabasz index, serta dilengkapi dengan visualisasi PCA dan pemetaan spasial. Dengan pendekatan ini, penelitian diharapkan dapat memberikan pemetaan risiko gempa bumi yang lebih representatif dan mendukung strategi mitigasi bencana di Indonesia.

2. METODOLOGI PENELITIAN

2.1. Alur Penelitian

Penelitian ini terdiri dari beberapa tahapan yang dilakukan untuk membangun model *clustering* gempa bumi di Indonesia. Rangkaian proses dalam penelitian ini dapat dilihat secara visual pada Gambar 1.



Gambar 1. Alur Penelitian

Penelitian ini diawali dengan identifikasi masalah, yaitu menentukan kendala utama dalam mengelompokkan wilayah rawan gempa menggunakan metode *clustering*. Dilanjutkan dengan pengumpulan dan eksplorasi data gempa untuk melihat pola awal. Data kemudian diproses melalui tahap *preprocessing* guna memastikan kualitas dan konsistensi.

Selanjutnya dilakukan klasterisasi menggunakan algoritma BIRCH dan DBSCAN untuk mengelompokkan wilayah berdasarkan frekuensi, rata-rata magnitudo, dan rata-rata kedalaman gempa. Hasil *cluster* dianalisis dan divisualisasikan, kemudian ditarik kesimpulan.

2.2. Pengumpulan Data

Data dalam penelitian ini diperoleh dari *European-Mediterranean Seismological Centre* (EMSC), sebuah lembaga internasional yang menyediakan informasi tentang kejadian gempa bumi dari berbagai wilayah di dunia secara *real-time*. *Dataset* yang digunakan terdiri dari 6.654 baris data yang mencakup kejadian gempa bumi terbaru hingga tahun 2025, di mana sampel data tersebut dapat dilihat pada Tabel I.

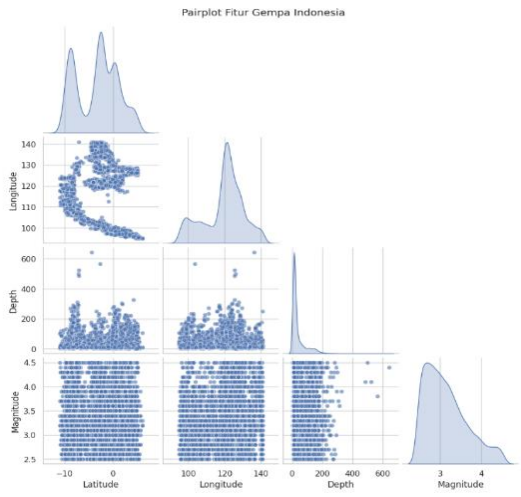
Setiap baris dalam *dataset* merepresentasikan satu kejadian gempa dan memuat informasi penting seperti tanggal (*Date*), waktu dalam format *Time* (UTC), koordinat lokasi (*Latitude* dan *Longitude*), kedalaman gempa (*Depth*), jenis magnitudo (*Magnitude Type*), nilai magnitudo (*Magnitude*), nama wilayah kejadian (*Region name*), serta identitas unik gempa (*EqId*). Data ini dihimpun dalam format terstruktur dan digunakan sebagai dasar dalam penerapan metode *clustering,* dengan tujuan untuk mengelompokkan kejadian gempa berdasarkan kemiripan karakteristiknya.

TABEL I. *DATASET*

Date	***Time (UTC)***	***Latitude***	***Longitude***	***Region name***	***Depth***	***Magnitude Type***	***Magnitude***	***Eqld***
7/8/2025	17:02:42	-8.64	118.41	SUMBAWA *REGION*, INDONESIA	98.0	m	4.1	20250708_0000178
7/8/2025	16:37:51	-1.02	137.25	*NEAR N COAST OF* PAPUA, INDONESIA	98.0	m	4.4	20250708_0000174
7/8/2025	14:10:37	-2.20	127.71	*CERAM SEA*, INDONESIA	13.0	m	3.7	20250708_0000144
7/8/2025	13:33:02	-5.61	121.81	SULAWESI, INDONESIA	10.0	m	2.8	20250708_0000134
7/8/2025	12:22:55	2.60	128.32	HALMAHERA, INDONESIA	16.0	m	3.3	20250708_0000122

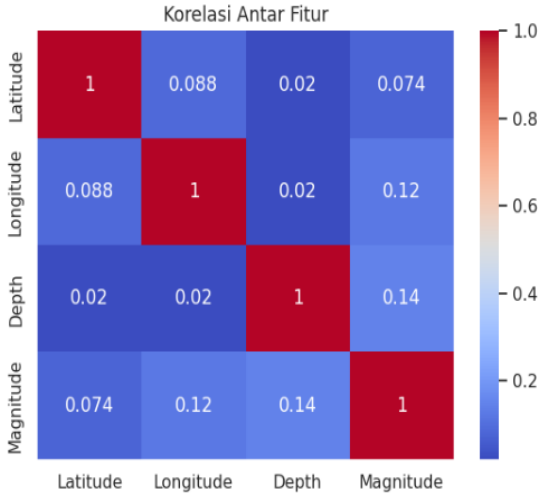
2.3. Eksplorasi Data

Eksplorasi data adalah proses analisis awal yang bertujuan untuk memperoleh pemahaman yang lebih menyeluruh mengenai karakteristik dan pola yang terkandung dalam data [8]. Visualisasi awal menggunakan *pairplot* dimanfaatkan untuk melihat distribusi setiap variabel serta keterkaitan antar fitur utama seperti *Latitude*, *Longitude*, *Depth*, dan *Magnitude* sebagaimana divisualisasikan pada Gambar 2.



Gambar 2. *Pairplot*

Melalui visualisasi ini, indikasi awal terhadap kemungkinan terbentuknya *cluster* atau pola tersembunyi dalam data dapat diamati secara lebih menyeluruh.



Gambar 3. *Heatmap*

Selain itu, Gambar 3 memperlihatkan visualisasi *heatmap* korelasi yang digunakan untuk menganalisis keterkaitan antar variabel numerik berdasarkan nilai korelasinya. Korelasi antar fitur ini penting untuk memastikan bahwa tidak terdapat redundansi informasi yang berlebihan yang dapat memengaruhi proses klasterisasi. Eksplorasi ini berperan sebagai landasan awal dalam memahami struktur data sebelum dilanjutkan ke tahap *preprocessing*.

2.4. Preprocessing Data

Preprocessing merupakan langkah yang sangat penting dalam meningkatkan kualitas data dan performa model [9]. Pada tahap awal *preprocessing*, dilakukan penghapusan terhadap nilai kosong (*missing values*) dengan tujuan untuk menghindari gangguan dalam proses *data mining*, khususnya *clustering,* sehingga data yang

digunakan menjadi lebih bersih dan siap dianalisis secara optimal [10].

Selanjutnya, dilakukan proses penyaringan (*filtering*) untuk menghilangkan data yang tidak relevan. *Filtering* yang dilakukan mencakup penyaringan berdasarkan wilayah, yaitu dengan memilih baris data yang mengandung kata 'Indonesia' pada atribut *Region name*, agar analisis difokuskan hanya pada kejadian gempa yang terjadi di wilayah Indonesia. Setelah melewati tahap ini, *dataset* berkurang menjadi 5.516 baris, yang seluruhnya merepresentasikan kejadian gempa bumi di wilayah Indonesia.

Tahap selanjutnya adalah transformasi data ke dalam bentuk *grid* spasial, di mana wilayah Indonesia dibagi menjadi sel-sel *grid*. Pada setiap *grid*, dihitung atribut yang merepresentasikan karakteristik wilayah berdasarkan kejadian gempa yang terjadi di dalamnya, meliputi frekuensi kejadian gempa, rata-rata magnitudo, dan rata-rata kedalaman. Atribut-atribut ini digunakan sebagai fitur utama dalam proses analisis klasterisasi. Ketiga atribut ini kemudian dinormalisasi menggunakan metode *Z-Score* dengan bantuan *StandardScaler* untuk memastikan bahwa setiap fitur memiliki skala yang setara. Normalisasi ini penting dilakukan agar tidak terjadi dominasi fitur tertentu dalam proses klasterisasi [11].

2.5. Clustering

Clustering adalah proses mengelompokkan data atau observasi ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Berbeda dengan klasifikasi, *clustering* tidak menggunakan variabel target sebagai acuan. Tujuan utamanya bukan untuk memprediksi atau mengestimasi nilai tertentu, melainkan untuk membagi data secara alami menjadi kelompok-kelompok yang anggotanya memiliki kesamaan satu sama lain. Dengan begitu, pola tersembunyi dalam data dapat terungkap tanpa harus mengetahui hasil akhir sejak awal [12][13].

2.6. BIRCH *Clustering*

*Balanced Iterative Reducing and Clustering using Hierarchies (*BIRCH) merupakan algoritma klasterisasi yang dirancang untuk menangani data berukuran besar secara efisien. BIRCH memanfaatkan struktur *Clustering Feature Tree* (*CF Tree*) untuk merangkum data secara hierarkis dan memungkinkan pemrosesan data secara inkremental dalam satu atau beberapa kali pemindaian.

Inti dari algoritma ini adalah *Clustering Feature* (CF), yaitu representasi ringkas dari sekelompok data yang dirumuskan sebagai berikut:

$$CF = (N, LS, SS) \quad (1)$$

Keterangan:
N: jumlah data dalam *cluster*
$LS = \sum p_i$: jumlah total vektor data
$SS = \sum |p_i|^2$: jumlah kuadrat norma vektor data

Dua *Clustering Feature* dapat digabungkan menggunakan prinsip aditif:

$$CFgabung = (N_1 + N_2, LS_1 + LS_2, SS_1 + SS_2) \quad (2)$$

Proses pembentukan *CF Tree* dilakukan dengan membandingkan jarak *Euclidean* antar node, dan data baru akan ditempatkan ke node terdekat selama masih berada dalam ambang batas radius tertentu. Jika syarat ini terpenuhi, struktur pohon akan diperbarui secara lokal tanpa harus memproses seluruh data ulang [14].

2.7. DBSCAN Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) merupakan algoritma *unsupervised learning* berbasis kepadatan yang banyak digunakan dalam analisis *cluster* data, khususnya ketika bentuk *cluster* tidak beraturan dan data mengandung *noise*. DBSCAN bekerja dengan prinsip utama bahwa sebuah *cluster* terdiri atas sekumpulan titik data yang saling terhubung secara kepadatan, tanpa memerlukan penentuan jumlah *cluster* di awal. Hal ini menjadi keunggulan tersendiri dibanding metode seperti K-Means, yang mengharuskan pengguna menentukan jumlah *cluster* terlebih dahulu [15].

Konsep dasar DBSCAN mencakup dua parameter utama, yaitu ε (*epsilon*) dan *MinPts* (*minimum points*). Parameter ε menentukan radius atau jarak maksimum antara dua titik agar dianggap bertetangga, sementara *MinPts* adalah jumlah minimum tetangga yang diperlukan agar sebuah titik diklasifikasikan sebagai *core point*. Titik-titik yang memiliki cukup tetangga dalam *radius* ε akan menjadi titik inti (*core*), sedangkan titik yang berada dalam *radius* ε dari *core point* namun tidak memiliki cukup tetangga sendiri disebut *border point*. Titik-titik yang tidak dapat dijangkau oleh titik inti mana pun akan dianggap sebagai *noise* atau *outlier* [16]. Jarak antar titik

biasanya dihitung menggunakan jarak *Euclidean*, yang dirumuskan sebagai berikut:

$$d(P,C) = \sqrt{\sum_{i=1}^{n} (x_{pi} - x_{ci})^2} \qquad (3)$$

Keterangan:
$d(P,C)$ = Jarak *Euclidean* antara titik data P dan C (biasanya pusat *cluster*)
x_{pi} = Nilai fitur ke-i dari titik P
x_{ci} = Nilai fitur ke-i dari titik C
n = Jumlah dimensi atau fitur data

2.8. Analisis dan Visualisasi

Pada tahap ini, dilakukan perbandingan hasil pengelompokan dari dua algoritma, yaitu DBSCAN dan BIRCH, untuk menilai sejauh mana masing-masing metode mampu membentuk *cluster* yang representatif terhadap pola data gempa. Penilaian kualitas *cluster* dilakukan menggunakan beberapa metrik evaluasi, yaitu *Silhouette coefficient*, *Davies-Bouldin index*, dan *Calinski-Harabasz index*. *Silhouette coefficient* digunakan untuk mengukur keserupaan data dalam satu *cluster* dibandingkan dengan *cluster* lain [17], sedangkan *Davies-Bouldin index* menilai seberapa baik pemisahan antar cluster dilakukan. Semakin rendah nilainya, semakin baik kualitas *cluster* [18]. Sementara itu, *Calinski-Harabasz index* mengevaluasi rasio antara variansi antar *cluster* dan variansi dalam *cluster*, di mana nilai yang lebih tinggi menunjukkan pemisahan *cluster* yang lebih baik [19].

Setelah evaluasi, masing-masing *cluster* dianalisis untuk mengidentifikasi potensi risikonya. Tingkat risiko ditentukan berdasarkan tiga aspek utama, yaitu frekuensi kejadian, magnitudo, dan kedalaman gempa. *Cluster* dengan frekuensi tinggi, magnitudo besar, dan kedalaman dangkal dikategorikan sebagai Rawan, sedangkan *cluster* lainnya sebagai Tidak Rawan. Hasil analisis ini divisualisasikan dalam bentuk grafik dan peta sebaran untuk memberikan gambaran yang lebih jelas terkait pola dan potensi bahaya gempa di Indonesia.

2.9. *Google Colaboratory*

Google Colaboratory (Colab) merupakan platform komputasi awan yang disediakan oleh *Google* untuk memfasilitasi kegiatan analisis data dan penelitian di bidang *machine learning* [20]. Platform ini berbasis pada *Jupyter Notebook*, namun terintegrasi dengan layanan *Google* sehingga memungkinkan pengguna untuk menulis, menjalankan, serta membagikan kode pemrograman secara kolaboratif melalui peramban *web*.

Dalam penelitian ini, penerapan algoritma *data mining* untuk proses *clustering* data gempa bumi di Indonesia dilakukan sepenuhnya melalui platform *Google Colaboratory*. Pemilihan Colab didasarkan pada kestabilan layanan berbasis *cloud*, kemudahan integrasi dengan *Google Drive* untuk pengelolaan *dataset*, serta efisiensinya dalam menangani proses komputasi intensif, khususnya saat pembentukan dan evaluasi model *clustering*. Selain itu, dukungan pustaka ilmiah seperti *Scikit-learn*, *Pandas*, dan *Matplotlib* yang telah tersedia secara default memungkinkan proses analisis data dilakukan secara cepat dan terstruktur tanpa perlu konfigurasi tambahan.

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data

Data yang digunakan dalam penelitian ini diperoleh dari situs resmi *European-Mediterranean Seismological Centre* (EMSC), "*www.emsc-csem.org/Earthquake_information*". Dataset mencakup 6.654 kejadian gempa bumi yang terjadi di wilayah Indonesia hingga tahun 2025. Setiap baris dalam dataset memuat informasi seperti tanggal, waktu, koordinat lokasi, kedalaman, magnitudo, dan nama wilayah kejadian gempa.

Eksplorasi awal dilakukan untuk memahami karakteristik data secara menyeluruh. Visualisasi menggunakan *pairplot* menunjukkan sebaran serta pola hubungan antar fitur utama seperti *Latitude*, *Longitude*, *Depth*, dan *Magnitude*. Selain itu, analisis korelasi antar variabel numerik ditampilkan dalam bentuk *heatmap* untuk memastikan tidak terdapat fitur yang tumpang tindih informasinya, yang dapat memengaruhi efektivitas proses klasterisasi.

Setelah itu, dilakukan tahap *preprocessing* data dengan menghapus nilai kosong (*missing values*) guna menjaga kualitas informasi yang dianalisis. Data kemudian difilter dengan memilih baris yang mengandung kata "Indonesia" pada atribut *Region name*, sehingga diperoleh sebanyak 5.516 baris data yang relevan dengan wilayah penelitian.

Tahap selanjutnya adalah transformasi data ke dalam *grid* spasial, di mana wilayah Indonesia dibagi berdasarkan koordinat geografis. Setiap *grid* dihitung berdasarkan frekuensi gempa, rata-rata magnitudo, dan rata-rata kedalaman sebagai

fitur utama klasterisasi, sebagaimana ditunjukkan melalui beberapa baris hasil penghitungan pada Tabel II. Seluruh fitur kemudian dinormalisasi menggunakan *Z-Score* agar memiliki skala yang setara dan tidak saling mendominasi dalam proses klasterisasi.

TABEL II. STATISTIK GEMPA BERDASARKAN *GRID*

Lat_ Bin	***Lon_ Bin***	***Frequency***	***Mean_ Magnitude***	***Mean_ Depth***
-11.0	117.8	1	3.7	10.0
-10.9	114.4	1	4.5	10.0
-10.9	117.7	1	3.4	5.0
-10.9	123.6	1	3.2	104.0
-10.8	110.9	1	4.4	23.0

3.2. BIRCH *Clustering*

Dalam penelitian ini, BIRCH dipilih karena efisien dalam menangani dataset besar dan mampu membentuk *cluster* secara hierarkis tanpa banyak iterasi atau ketergantungan pada parameter awal. Algoritma ini cocok untuk data spasial gempa bumi yang memiliki sebaran bervariasi. Hasil evaluasi performa BIRCH untuk jumlah *cluster* k = 2 hingga 5 disajikan pada Tabel III berikut.

TABEL III. HASIL EVALUASI BIRCH

Jumlah *cluster*	***Silhouette coefficient***	***Davies-Bouldin index***	***Calinski-Harabasz index***
2	0.8910	0.3580	1009.83
3	0.6008	0.5542	899.02
4	0.6004	0.6137	646.28
5	0.5604	0.6035	555.69

Berdasarkan Tabel III, konfigurasi jumlah *cluster* k = 2 menunjukkan performa terbaik dibandingkan nilai k lainnya. Hal ini ditunjukkan oleh nilai *Silhouette coefficient* tertinggi sebesar 0.8910, yang mengindikasikan pemisahan *cluster* yang sangat baik, serta *Davies-bouldin Index* terendah sebesar 0.3580, yang menunjukkan bahwa *cluster* yang terbentuk memiliki jarak antar *cluster* yang jelas dan kompak secara internal. Selain itu, nilai *Calinski-Harabasz index* tertinggi sebesar 1009.83 memperkuat bahwa struktur *cluster* pada k = 2 paling optimal dalam memisahkan data.

3.3. DBSCAN *Clustering*

Setelah penerapan BIRCH, algoritma selanjutnya yang digunakan adalah DBSCAN, yang memiliki pendekatan berbeda dalam menentukan struktur *cluster*, yakni berdasarkan kepadatan titik data. Tidak seperti algoritma klasterisasi berbasis partisi seperti BIRCH, DBSCAN tidak memerlukan penentuan jumlah *cluster* di awal dan mampu mengidentifikasi *cluster* dengan bentuk yang tidak beraturan serta mengelompokkan data *outlier* sebagai *noise*. Hasil evaluasi performa DBSCAN dengan berbagai nilai parameter *epsilon* (ε) disajikan pada Tabel IV berikut.

TABEL IV. HASIL EVALUASI DBSCAN

Epsilon	**Jumlah *cluster***	***Jumlah Noise***	***Silhouette coefficient***	***Davies-Bouldin index***	***Calinski-Harabasz index***
2.70	2	9 (0.28%)	0.8120	0.3352	643.62
2.90	2	9 (0.28%)	0.8120	0.3352	643.62
3.10	2	4 (0.12%)	0.8850	4.2165	506.31
3.30	2	2 (0.06%)	0.8726	0.5017	554.37

Berdasarkan Tabel IV, konfigurasi *epsilon* sebesar 3.30 menghasilkan *Silhouette coefficient* sebesar 0.8726, yang menunjukkan kualitas pemisahan *cluster* yang sangat baik. Selain itu, *Davies-Bouldin index* sebesar 0.5017 menandakan *cluster* yang terbentuk memiliki kepadatan internal yang cukup baik dan jarak antar *cluster* yang cukup jelas. *Calinski-Harabasz index* sebesar 554.37 berada pada tingkat menengah dibanding konfigurasi lainnya, yang menunjukkan keseimbangan antara pemisahan *cluster* dan kekompakan internal. Dengan jumlah *noise* yang sangat kecil, yaitu hanya 0.06%, konfigurasi ini dianggap paling stabil dan seimbang dalam membentuk dua *cluster* utama tanpa mengorbankan terlalu banyak data.

3.4. Analisis dan Visualisasi

Analisis awal dilakukan dengan membandingkan hasil evaluasi performa terbaik antara algoritma BIRCH dan DBSCAN. Perbandingan ini mencakup tiga metrik utama, yaitu *Silhouette coefficient*, *Davies-Bouldin index*, dan *Calinski-Harabasz index*. Hasil lengkap perbandingan tersebut ditampilkan pada Tabel V.

TABEL V. PERBANDINGAN HASIL EVALUASI BIRCH DAN DBSCAN

Model	**Jumlah *cluster***	***Silhouette coefficient***	***Davies-Bouldin index***	***Calinski-Harabasz index***
BIRCH (k=2)	2	0.8910	0.3580	1009.83
DBSCAN (ε=3.30)	2	0.8726	0.5017	554.37

Berdasarkan Tabel V, model BIRCH dengan k = 2 menunjukkan performa terbaik secara keseluruhan, dengan nilai *Silhouette coefficient* tertinggi, *Davies-Bouldin index* terendah, dan *Calinski-Harabasz index* tertinggi, yang menandakan kualitas klasterisasi yang sangat baik dari segi pemisahan dan kekompakan. Sementara itu, DBSCAN dengan ε = 3.30 menghasilkan hasil yang kompetitif, dengan *Silhouette coefficient* dan *Davies-Bouldin index* yang masih tergolong baik, serta jumlah noise yang sangat rendah. Namun, secara umum performanya masih berada di bawah model BIRCH.

Sebagai bagian dari analisis lanjutan, dilakukan perbandingan statistik masing-masing *cluster* yang dihasilkan oleh BIRCH dan DBSCAN berdasarkan rata-rata magnitudo, kedalaman, dan jumlah kejadian gempa. Hasilnya disajikan pada Tabel VI.

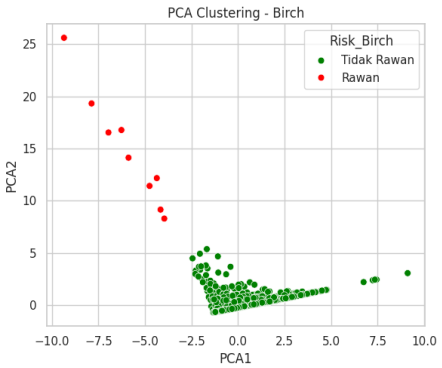
TABEL VI. PERBANDINGAN STATISTIK *CLUSTER* BIRCH DAN DBSCAN

Model	***Cluster***	***Mean_ Magnitude***	***Mean_ Depth***	***Frequency***
BIRCH	0	3.16	40.42	5009
BIRCH	1	3.04	9.65	507
DBSCAN	0	3.16	40.42	5009
DBSCAN	1	3.07	9.62	270

Berdasarkan Tabel VI, baik BIRCH maupun DBSCAN menghasilkan dua *cluster* utama dengan pola yang serupa. *Cluster* 0 pada keduanya mencakup sebagian besar data, dengan rata-rata magnitudo 3.16 dan kedalaman 40.42 km, mencerminkan gempa menengah hingga dalam.

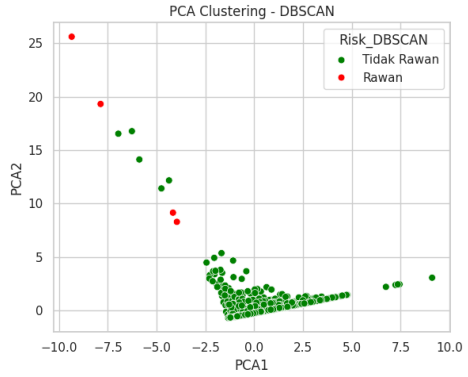
Cluster 1 menunjukkan karakteristik gempa dangkal, dengan kedalaman sekitar 9.6 km. BIRCH mengelompokkan 507 kejadian, sementara DBSCAN lebih selektif dengan hanya 270 kejadian, meskipun memiliki magnitudo rata-rata sedikit lebih tinggi, yaitu 3.07. Hal ini menunjukkan bahwa DBSCAN cenderung lebih konservatif dalam membentuk *cluster* berbasis kepadatan, sedangkan BIRCH lebih inklusif terhadap variasi data.

Sebagai upaya untuk mempermudah interpretasi hasil klasterisasi dalam ruang dua dimensi, digunakan teknik *Principal Component Analysis* (PCA). Teknik ini mereduksi dimensi dari fitur numerik seperti frekuensi kejadian, rata-rata magnitudo, dan rata-rata kedalaman menjadi dua komponen utama yang merepresentasikan sebagian besar variasi dalam data.



Gambar 4. Visualisasi PCA *Cluster* BIRCH

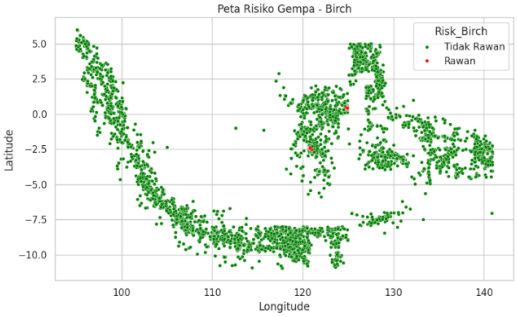
Gambar 4 menunjukkan visualisasi hasil klasterisasi menggunakan algoritma BIRCH dalam ruang dua dimensi hasil reduksi PCA. Titik berwarna hijau merepresentasikan wilayah tidak rawan, sedangkan titik merah menunjukkan wilayah rawan gempa. Terlihat bahwa mayoritas data terkonsentrasi pada *cluster* tidak rawan, sementara *cluster* rawan terpisah cukup jelas, menandakan bahwa BIRCH mampu membedakan karakteristik wilayah berdasarkan data gempa dengan baik.



Gambar 5. Visualisasi PCA *Cluster* DBSCAN

Gambar 5 merupakan visualisasi hasil klasterisasi menggunakan algoritma DBSCAN dalam ruang dua dimensi setelah direduksi menggunakan PCA. Titik hijau menunjukkan wilayah yang diklasifikasikan sebagai tidak rawan, sedangkan titik merah menunjukkan wilayah rawan gempa. Dibandingkan BIRCH, DBSCAN menghasilkan lebih sedikit wilayah rawan dan distribusinya lebih tersebar, mencerminkan pendekatannya yang selektif namun cenderung kurang sensitif dalam mengenali pola *cluster* yang lebih luas, terutama pada data dengan sebaran tidak merata.

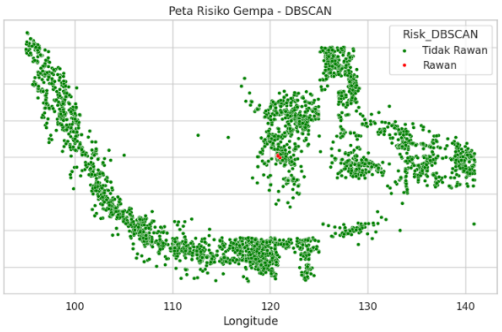
Selain visualisasi PCA, dilakukan pemetaan spasial hasil klasterisasi untuk melihat sebaran geografis *cluster* berdasarkan wilayah di Indonesia. Pemetaan ini membantu mengidentifikasi pola kerawanan gempa secara langsung pada konteks geografis.



Gambar 6. Peta Sebaran *Cluster* BIRCH

Gambar 6 memperlihatkan peta sebaran hasil klasterisasi menggunakan algoritma BIRCH berdasarkan koordinat geografis wilayah Indonesia. Titik hijau menunjukkan lokasi yang diklasifikasikan sebagai tidak rawan gempa, sedangkan titik merah menandai wilayah yang termasuk dalam *cluster* rawan gempa.

Sebaran *cluster* rawan terlihat terkonsentrasi di beberapa area seperti bagian tengah dan timur Indonesia, yang secara geologis memang berada di zona subduksi aktif. Sementara itu, *cluster* tidak rawan tersebar luas di hampir seluruh wilayah, menunjukkan bahwa sebagian besar lokasi memiliki karakteristik gempa dengan kedalaman atau intensitas yang lebih rendah.



Gambar 7. Peta Sebaran *Cluster* DBSCAN

Gambar 7 memperlihatkan peta sebaran hasil klasterisasi menggunakan algoritma DBSCAN berdasarkan lokasi geografis gempa di Indonesia. Titik berwarna hijau menunjukkan wilayah yang dikategorikan sebagai tidak rawan, sedangkan titik merah menandai wilayah yang teridentifikasi sebagai rawan gempa.

Dibandingkan BIRCH, DBSCAN menghasilkan wilayah rawan yang lebih sedikit dan terlokalisasi. Meskipun lebih selektif, pendekatan ini cenderung kurang sensitif terhadap pola *cluster* yang lebih luas, sehingga berisiko mengabaikan beberapa wilayah berisiko.

4. KESIMPULAN DAN SARAN

Hasil pengujian menunjukkan bahwa algoritma BIRCH berkinerja terbaik, dengan koefisien Silhouette sebesar 0,8910, indeks Davies-Bouldin sebesar 0,3580, dan indeks Calinski-Harabasz sebesar 1009,83. Angka-angka ini menunjukkan bahwa BIRCH mampu membentuk klaster yang lebih jelas, lebih padat, dan lebih stabil. Sementara itu, DBSCAN, dengan parameter optimal, menghasilkan hasil yang relatif baik (koefisien Silhouette sebesar 0,8726, indeks Davies-Bouldin sebesar 0,5017, dan indeks Calinski-Harabasz sebesar 554,37), tetapi masih berkinerja di bawah BIRCH, terutama dalam hal konsistensi pengelompokan.

Temuan ini menunjukkan bahwa studi ini memberikan kontribusi penting, yang menunjukkan bahwa BIRCH lebih efektif untuk pemetaan risiko gempa bumi di Indonesia, terutama karena kemampuannya menangani kumpulan data besar dengan variasi kedalaman yang kompleks. Hal ini membuka peluang untuk penerapan praktis di lapangan, misalnya, dengan mendukung lembaga penanggulangan bencana

dan perencana daerah dalam mengembangkan prioritas mitigasi yang lebih terarah di daerah rawan gempa bumi.

Penelitian di masa mendatang dapat berfokus pada penambahan atribut geologis seperti jenis tanah, jarak ke patahan aktif, dan kedalaman lempeng subduksi untuk lebih menyelaraskan hasil pemetaan dengan kondisi dunia nyata. Lebih lanjut, pendekatan hibrida yang menggabungkan keunggulan BIRCH dan DBSCAN juga berpotensi menghasilkan model yang lebih adaptif untuk mengenali pola gempa bumi dinamis dan mendeteksi anomali dengan lebih baik.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyusunan penelitian ini. Secara khusus, penulis menyampaikan terima kasih kepada Edwinsyah Umasugi, S.Kom atas bimbingan, arahan, serta motivasi yang diberikan selama proses penelitian berlangsung. Penulis juga berterima kasih kepada *European-Mediterranean Seismological Centre* (EMSC) atas tersedianya data gempa bumi yang menjadi dasar analisis dalam penelitian ini. Semoga hasil penelitian ini dapat memberikan manfaat dalam pengembangan studi mengenai gempa bumi dan peningkatan kewaspadaan terhadap risiko bencana di Indonesia.

Daftar Pustaka:

[1] A. Prasetio, M. M. Effendi, and M. N. Dwi M, "Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering," *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 338–343, 2023, doi: 10.47065/bit.v4i3.820.

[2] M. Aviedo Murel, M. Febri Yoga Saputra, E. Kristian, F. Andrea Micelle, and N. Kristianti, "Analisis Kawasan Rawan Bencana Gempa Bumi Di Aceh, Yogyakarta, Dan Sulawesi Tengah Menggunakan Metode Polygon Pada Aplikasi Qgis," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 4194–4199, 2024, doi: 10.36040/jati.v8i3.9882.

[3] B. Putra Aryadi and N. Hendrastuty, "Penerapan Algoritma K-Means Untuk Melakukan Klasterisasi Pada Varietas Padi," *J. Inform. Rekayasa Elektron.*, vol. 7, no. 1, pp. 124–129, 2024, [Online]. Available: http://e-journal.stmiklombok.ac.id/index.php/jir eISSN.2620-6900

[4] I. H. Rifa, H. Pratiwi, and R. Respatiwulan, "Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms," *Media Stat.*, vol. 13, no. 2, pp. 194–205, 2020, doi: 10.14710/medstat.13.2.194-205.

[5] A. P. W. Hadi, H. Pratiwi, and I. Slamet, "Pengelompokan Data Gempa Bumi di Indonesia dengan Algoritma K-Means dan DBSCAN," *Semin. Nas. Pendidik.*, pp. 52–60, 2023, [Online]. Available: http://seminar.uad.ac.id/index.php/sendikmad/article/view/12512

[6] J. Inayah, A. Fanani, and W. D. Utami, "Klasterisasi Data Kejadian Gempa Bumi di Indonesia Menggunakan Metode K-Medoids," *J. Sist. dan Teknol. Inf.*, vol. 12, no. 2, p. 271, 2024, doi: 10.26418/justin.v12i2.73594.

[7] F. Reviantika, C. N. Harahap, and Y. Azhar, "Analisis Gempa Bumi pada Pulau Jawa menggunakan Clustering Algoritma K-Means," *J. Din. Inform.*, vol. 9, no. 1, pp. 51–60, 2020, [Online]. Available: https://twitter.com/infobmkg

[8] I. N. Rizki, D. Prayoga, M. L. Puspita, and M. Q. Huda, "Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3856.

[9] J. E. Widyaya and S. Budi, "Pengaruh Preprocessing Terhadap Klasifikasi Diabetic Retinopathy dengan Pendekatan Transfer Learning Convolutional Neural Network," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 110–124, 2021, doi: 10.28932/jutisi.v7i1.3327.

[10] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[11] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.

[12] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk

Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1–7, 2022.

[13] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.

[14] F. Artamevia, P. Ristiawan, A. Purno, and W. Wibowo, "Analysis and Clustering of Poverty Levels by Education in Cimahi City Using the BIRCH Method (Balanced Iterative Reducing and Clustering using Hierarchies)," *Brill. Res. Artif. Intell.*, vol. 5, no. 1, pp. 12–20, 2025.

[15] I. N. Simbolon and P. D. Friskila, "Analisis Dan Evaluasi Algoritma Dbscan (Density-Based Spatial Clustering of Applications With Noise) Pada Tuberkulosis," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3s1.5206.

[16] A. S. Devi, I. K. G. D. Putra, and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 6, no. 3, p. 185, 2015, doi: 10.24843/lkjiti.2015.v06.i03.p05.

[17] Y. Hasan, "Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Cluster K-Means Dan Dbscan," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, pp. 60–74, 2024, doi: 10.23960/jitet.v12i3s1.5001.

[18] I. F. Fauzi, M. G. Resmi, and T. I. Hermanto, "Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit," *JUMANJI (Jurnal Masy. Inform. Unjani)*, vol. 7, no. 2, pp. 1–15, 2023, [Online]. Available: https://jumanji.unjani.ac.id/index.php/jumanji/article/view/321

[19] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 89–97, 2023, doi: 10.30871/jaic.v7i1.4947.

[20] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.