

# Pengklasteran Kejadian Gempa Wilayah Indonesia Menggunakan Algoritma k-Means

I Made Budi Suksmadana<sup>1</sup>, Cipta Ramadhani<sup>1</sup>

<sup>1</sup>Jurusan Teknik Elektro, Fakultas Teknik, Universitas Mataram, Jln. Majapahit No. 62, Mataram 83125, Indonesia

## ARTICLE INFO

## ABSTRACT

### Article history:

Received January 29, 2022

Revised February 16, 2022

Accepted February 28, 2022

### Keywords:

earthquake;

k-Means;

plate;

machine learning;

*Earthquakes can cause human casualties due to the collapse of buildings and also cause secondary hazards such as fires, landslides and tsunamis. Indonesia is an area that often experiences earthquakes because it is surrounded by four main plates, namely the Eurasian Plate, the Indo-Australian Plate, the Philippine Sea Plate, and the Pacific Plate. Earthquake events in an area can be clustered into certain groups. With these clusters, it is hoped that one can find out the relationship between one earthquake event and another. The k-Means method is a clustering method that is part of Scikit-Learn, a machine learning library. The k-Means method is carried out using the values of  $k = 3, 5$  and  $7$ . Clustering with  $k = 7$  produces a maximum number of earthquake events of 571 in one cluster and a minimum of 180 in another cluster.*

Gempa bumi dapat menyebabkan jatuhnya korban jiwa manusia yang disebabkan karena runtuhnya bangunan dan juga menyebabkan bahaya sekunder seperti kebakaran, tanah longsor dan tsunami. Indonesia merupakan wilayah yang sering mengalami kejadian gempa karena dikelilingi oleh empat lempeng utama, yaitu Lempeng Eurasia, Lempeng Indo-Australia, Lempeng Laut Filipina, dan Lempeng Pasifik. Kejadian gempa dalam sebuah wilayah dapat diklaster menjadi kelompok-kelompok tertentu. Dengan klaster-klaster tersebut diharapkan dapat mengetahui hubungan satu kejadian gempa dengan gempa yang lainnya. Metode *k-Means* adalah salah satu metode pengklasteran yang merupakan bagian dari Scikit-Learn, yaitu sebuah pustaka machine learning. Metode *k-Means* yang dilakukan menggunakan nilai  $k = 3, 5$  dan  $7$ . Pengklasteran dengan  $k = 7$  menghasilkan jumlah kejadian gempa maksimum sebanyak 571 pada sebuah klaster dan minimum sebanyak 180 pada sebuah klaster lainnya.

### Corresponding Author:

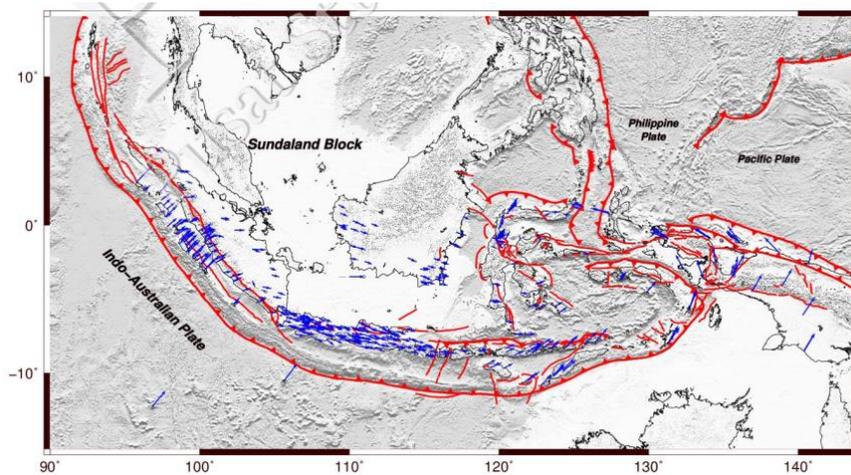
I Made Budi Suksmadana, Teknik Elektro Universitas Mataram, Jln. Majapahit No. 62, Mataram 83125, Indonesia

Email: mdbudisuk@unram.ac.id

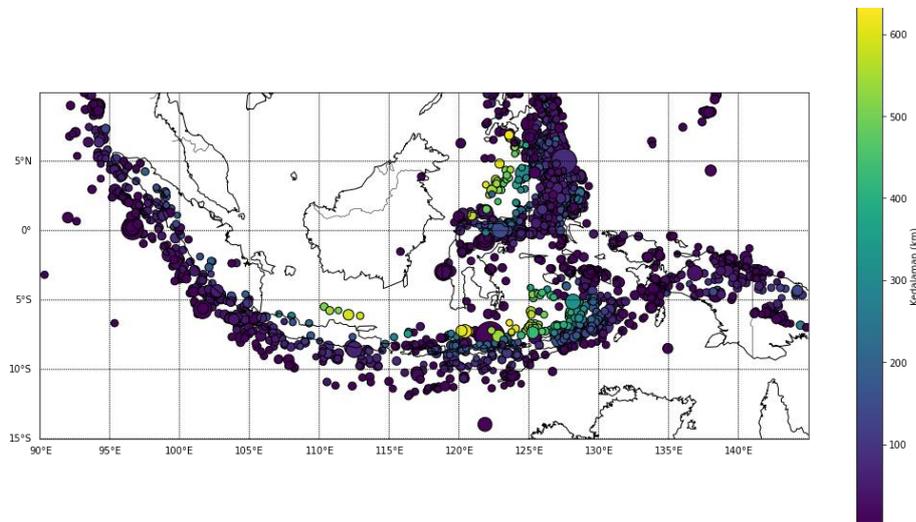
## 1. PENDAHULUAN

Gempa bumi dapat menyebabkan jatuhnya korban jiwa manusia yang disebabkan karena runtuhnya bangunan dan juga menyebabkan bahaya sekunder seperti kebakaran, tanah longsor dan tsunami. Gempa bumi dapat juga menyebabkan kerugian ekonomi yang besar dan merusak area yang luas. Indonesia merupakan wilayah yang sering mengalami kejadian gempa karena dikelilingi oleh empat lempeng utama, yaitu Lempeng Eurasia, Lempeng Indo-Australia, Lempeng Laut Filipina, dan Lempeng Pasifik “Gambar. 1”[1]. Beberapa kejadian gempa adalah rangkaian gempa Lombok yang terjadi beruntun (29 Juli 2018 (M6.4), 5 Agustus 2018 (M7.0) dan 19 Agustus 2018 (M6.9))[2], serta kejadian gempa Palu yang disertai tsunami dan likuifaksi (28 September 2018 (M7.4))[3] yang banyak menelan korban jiwa dan kerugian ekonomi yang besar dan merusak area yang luas.

Pengklasteran pernah diteliti dengan menggunakan algoritma DBSCAN dengan hasil wilayah resiko gempa dapat dikelompokkan menjadi 3 klaster dengan rincian, klaster 1 berjumlah 14 kejadian, klaster 2 berjumlah 1006 kejadian serta 8 kejadian untuk klaster 3[4]. Menurut data gempa yang diunduh dari United States Geological Survey (USGS)[5] terjadi sebanyak 2458 kejadian gempa pada wilayah Indonesia, yang ditunjukkan seperti pada **Gambar 2** pada sepanjang Tahun 2021.



**Gambar 1.** Peta tektonik wilayah Indonesia dari data geodetik hingga tahun 2016, vektor kecepatan pada referensi sistem ITRF 2008[1].



**Gambar 2.** Kejadian gempa tanggal 01 Januari 2021 sampai dengan 31 Desember 2021, magnitudo  $\geq 2.5$ , dan dengan batasan wilayah  $90^{\circ}\text{BT} - 145^{\circ}\text{BT}$  dan  $10^{\circ}\text{LU} - 15^{\circ}\text{LS}$

## 2. METODE PENELITIAN

### 2.1. Sumber Data.

Data yang digunakan pada penelitian ini adalah data kejadian gempa yang diunduh dari United States Geological Survey (USGS)[5] dengan batasan-batasan magnitudo sama dengan lebih besar dari 2.5, tanggal 01 Januari 2021 sampai dengan 31 Desember 2021 dan dengan batasan wilayah  $90^{\circ}\text{BT} - 145^{\circ}\text{BT}$  dan  $10^{\circ}\text{LU} - 15^{\circ}\text{LS}$  (**Gambar 2**). Data tersebut terdiri 2452 kejadian gempa. Dataset dari beberapa variabel yang akan digunakan, seperti dapat dilihat pada **Tabel 1**.

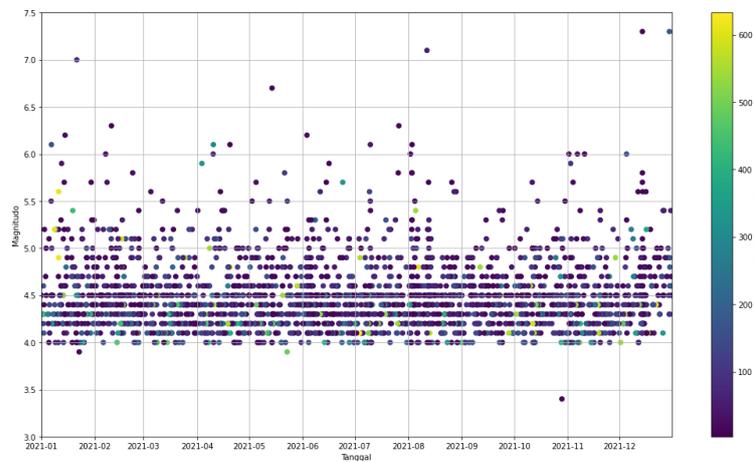
**Tabel 1.** Deskripsi variabel

Variabel	Tipe data (Python)
<i>time</i>	object
<i>latitude</i>	float64
<i>longitude</i>	float64
<i>depth</i>	float64
<i>mag</i>	float64

## 2.2. Pengolahan Data.

Pengolahan data dilakukan dengan menggunakan bahasa pemrograman Python, dengan menggunakan pustaka-pustaka “NumPy” (untuk operasi array), “pandas” (untuk membuat/manipulasi *dataframe*), “SciPy” (untuk statistik), “Matplotlib” (untuk visualisasi) dan “Scikit-Learn” (untuk pembelajaran mesin/*machine learning*)[6]. Langkah-langkah pengolahan data, yaitu:

1. Melakukan data *preprocessing* yang merupakan langkah awal terhadap *dataset* untuk mengetahui kelengkapan data dengan menggunakan “method info ()” dari pustaka “pandas”. Hasil yang didapat bahwa *dataset* terdiri dari 2458 baris dan 22 kolom dan setiap baris memiliki data yang lengkap.
2. Penseleksian *dataset* menggunakan pustaka “pandas” berdasarkan garis lintang dan bujur, tanggal awal dan tanggal akhir serta besaran magnitudo yang ditetapkan, sehingga jumlah *dataset* menjadi 2452 baris dan jumlah kolom tetap yaitu sebanyak 22 kolom.
3. Membuat visualisasi terhadap *dataset* yang sudah diseleksi pada langkah 2 (sesuai variabel-variabel pada Tabel 1.) dengan menggunakan pustaka “matplotlib” sehingga didapat hasil seperti pada **Gambar 2**.
4. Membuat visualisasi *dataset* antara waktu kejadian, magnitudo dan kedalaman gempa, sehingga didapat **Gambar 3**, yang nantinya digunakan untuk pertimbangan pengklasteran yang akan dirancang.



**Gambar 3.** Kejadian gempa berdasarkan waktu, magnitudo dan kedalaman gempa

5. Melakukan pengelompokan jumlah kejadian gempa berdasarkan rentang magnitudo dan rentang kedalaman, sehingga didapat **Tabel 2** dan **Tabel 3**.

**Tabel 2.** Jumlah kejadian gempa berdasarkan rentang magnitudo.

Rentang magnitudo	Jumlah gempa
$\leq 4$	104
$4 < - \leq 5$	2165
$5 < - \leq 6$	169
$> 6$	14

**Tabel 3.** Jumlah kejadian gempa berdasarkan rentang kedalaman.

Rentang kedalaman (km)	Jumlah gempa
$0 - \leq 33$	859
$33 < - \leq 70$	651
$70 < - \leq 150$	538
$150 < - \leq 300$	268
$300 < - \leq 500$	78
$> 500$	58

### 2.3. *k*-Means Clustering.

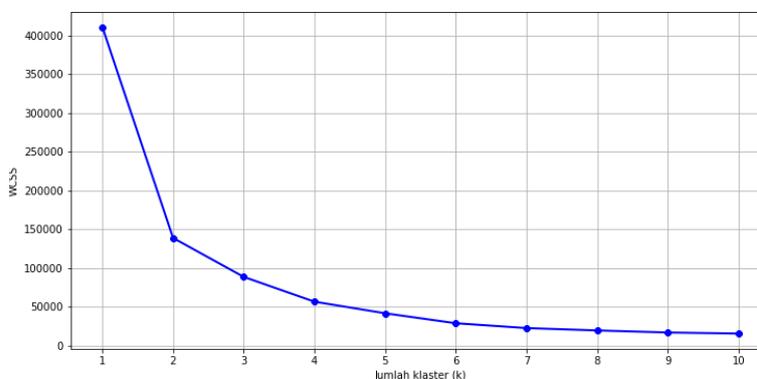
Metode *k*-Means clustering merupakan metode kuantisasi vektor yang awalnya berasal dari bidang pemrosesan sinyal dengan tujuan untuk memisahkan sebanyak  $n$  pengamatan menjadi  $k$  kluster/kelompok. Pada metode ini setiap pengamatan memiliki nilai rata-rata terdekat dengan pusat kluster (*cluster centroid*), sehingga menghasilkan kelompok-kelompok data. Metode *k*-Means clustering ini dapat meminimalkan varian dalam sebuah kluster (disebut jarak Euclidean kuadrat) tapi bukan jarak Euclidean seperti umumnya [7]. Algoritma *k*-Means umumnya menggunakan teknik perbaikan berulang yang umum disebut “*k*-Means algorithm” atau juga disebut “Lloyd algorithm” khususnya pada komunitas ilmu komputer. Kadang-kadang algoritma ini juga disebut sebagai “*naive k*-Means”, karena memiliki kemampuan yang lebih sederhana[8]. Langkah-langkah algoritma “*naive k*-Means” terdiri dari dua langkah berikut[9]:  
 Langkah penugasan: menugaskan setiap pengamatan ke cluster dengan mean terdekat yaitu yang memiliki jarak Euclidean kuadrat terkecil (secara matematis, ini berarti membagi pengamatan menurut diagram Voronoi yang dihasilkan), “(1)”.

$$S_i^{(t)} = \left\{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \quad \forall j, 1 \leq j \leq k \right\}, \quad (1)$$

dimana setiap  $x_p$  ditetapkan ke tepat satu  $S_i^{(t)}$  meskipun itu dapat ditetapkan ke dua atau lebih.  
 Langkah pembaharuan: hitung ulang nilai rata-rata ke pusat kluster (*centroid*) untuk pengamatan yang ditetapkan ke setiap kluster, “(2)”.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (2)$$

Algoritma dinyatakan telah konvergen ketika proses tidak lagi berubah. Algoritma ini tidak menjamin untuk menemukan nilai yang optimal[10]. Algoritma ini sering dinyatakan sebagai penugasan objek ke kluster terdekat berdasarkan jarak. Penggunaan fungsi jarak yang berbeda selain kuadrat jarak Euclidean dapat digunakan untuk mencapai konvergen. Berbagai modifikasi *k*-Means seperti *spherical k*-Means dan *k*-Medoids telah diusulkan untuk memungkinkan penggunaan fungsi jarak lainnya.

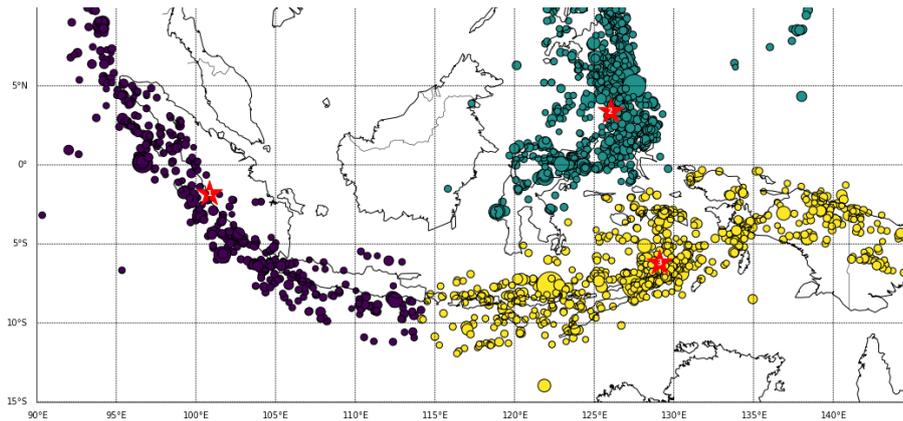


**Gambar 4.** Grafik Elbow untuk menentukan nilai  $k$  yang optimal

Pengklasteran yang ditawarkan menggunakan pembelajaran tipe tidak terbimbing (unsupervised learning) dengan algoritma *k*-Means, pembelajaran ini mencoba membuat kluster dan memasukkan data-data kedalamnya tanpa memberitahu kluster apa yang mungkin ada. Pembelajaran tidak terbimbing tidak memiliki target kluster, tetapi tidak berarti tidak dapat melakukan pengklasteran [11]. Anggota sebuah kluster serupa satu sama lainnya tetapi berbeda dengan anggota kluster lainnya. Pada bahasa Python pengklasteran dengan *k*-Means dilakukan dengan menggunakan “method KMeans” dari pustaka “sklearn.cluster”, penggunaan method ini harus menetapkan dahulu nilai  $k$  yang akan digunakan, nilai ini dapat ditentukan dengan menggunakan metode Within-Cluster-Sum-of-Squares (WCSS) untuk membuat grafik Elbow Gambar 4.

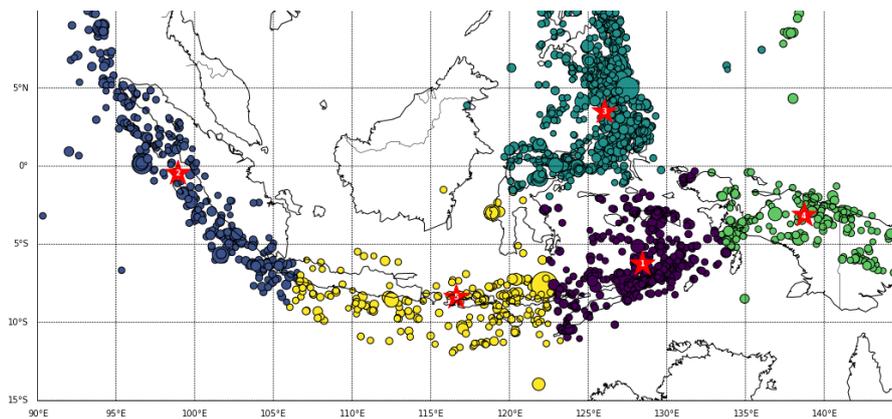
### 3. HASIL DAN PEMBAHASAN

Berdasarkan grafik Elbow **Gambar 4** untuk pengklasteran dapat dilakukan dengan nilai  $k > 2$ , dengan menetapkan nilai  $k$  bernilai ganjil yang diharapkan mendapatkan kluster yang proporsional. Jadi *dataset* diklaster dengan 3, 5, dan 7 kelompok, seperti terlihat pada **Gambar 5**, **Gambar 6**, dan **Gambar 7**.

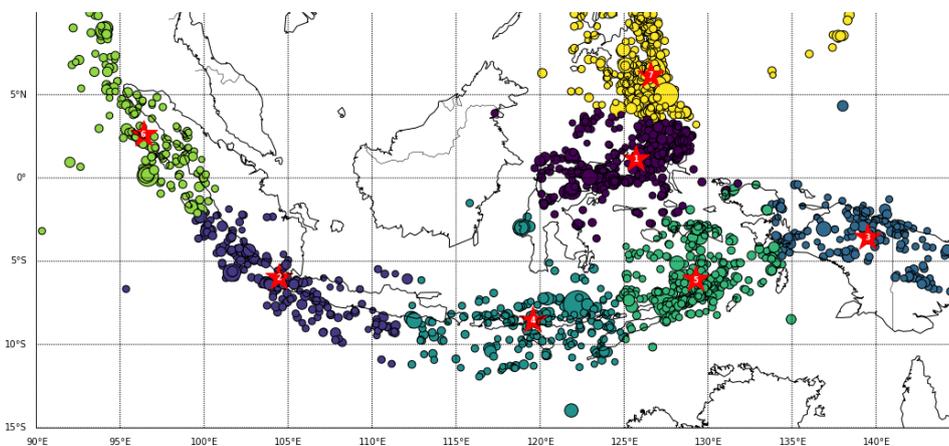


**Gambar 5.** Pengklasteran gempa dengan  $k = 3$ , (warna menunjukkan kluster, besar lingkaran menunjukkan magnitudo, bintang menunjukkan pusat kluster)

Pada pengklasteran dengan  $k = 3$  (**Gambar 5**), tampak kejadian gempa dikelompokkan menjadi dua bagian terlebih dahulu yaitu berdasarkan bujur, kemudian setengah bagian sebelah kanan dikelompokkan menjadi dua bagian berdasarkan lintang. Kejadian gempa kluster 1 sebanyak 476, kluster 2 sebanyak 1044 dan kluster 3 sebanyak 932.



**Gambar 6.** Pengklasteran gempa dengan  $k = 5$ , (warna menunjukkan kluster, besar lingkaran menunjukkan magnitudo, bintang menunjukkan pusat kluster)



**Gambar 7.** Pengklasteran gempa dengan  $k = 7$ , (warna menunjukkan kluster, besar lingkaran menunjukkan magnitudo, bintang menunjukkan pusat kluster)

**Gambar 6** merupakan hasil dari pengklasteran dengan  $k = 5$ , pada gambar ini tampak kejadian gempa diklaster menjadi 5 kelompok dengan masing-masing klaster jumlah kejadian gempanya yaitu: klaster 1 sebanyak 556, klaster 2 sebanyak 395, klaster 3 sebanyak 1012, klaster 4 sebanyak 216 dan klaster 5 sebanyak 273. Pemecahan klaster terjadi pada sebagian besar Pulau Jawa, Nusa Tenggara dan Papua, yang dipecah menjadi 3 klaster.

Hasil pengklasteran dengan  $k = 7$  ditampilkan pada **Gambar 7**, pada gambar ini tampak kejadian gempa diklaster menjadi 7 kelompok. Klaster 1 sebanyak 571, klaster 2 sebanyak 215, klaster 3 sebanyak 180, klaster 4 sebanyak 258, klaster 5 sebanyak 515, klaster 6 sebanyak 238 dan klaster 7 sebanyak 475. Pemecahan klaster terjadi pada Pulau Sumatra dan sebagian Pulau Jawa

#### 4. KESIMPULAN

Metode k-Means yang digunakan pada penelitian ini dapat digunakan untuk pengklasteran terhadap dataset kejadian gempa dalam setahun di Wilayah Indonesia. Metode k-Means dengan  $k = 7$  menghasilkan 7 buah klaster dengan jumlah kejadian gempa maksimum 571 dan minimum 180, gambar peklasteran ini sudah cukup rinci untuk seluruh kejadian gempa.

#### Ucapan Terima kasih

Terimakasih pada USGS yang telah mengizinkan untuk mengunduh dataset yang digunakan pada penelitian ini dan kepada pengembang pustaka-pustaka pada bahasa pemrograman Python.

#### REFERENCES

- [1] Tim Pusat Studi Gempa Nasional, PETA SUMBER DAN BAHAYA GEMPA INDONESIA TAHUN 2017. Pusat Penelitian dan Pengembangan Perumahan dan Permukiman Badan Penelitian dan Pengembangan Kementerian Pekerjaan Umum dan Perumahan Rakyat, 2017.
- [2] Tim Pusat Studi Gempa Nasional, KAJIAN RANGKAIAN GEMPA LOMBOK PROVINSI NUSA TENGGARA BARAT. Pusat Penelitian dan Pengembangan Perumahan dan Permukiman Badan Penelitian dan Pengembangan Kementerian Pekerjaan Umum dan Perumahan Rakyat, 2018.
- [3] Tim Pusat Studi Gempa Nasional, KAJIAN GEMPA PALU PROVINSI SULAWESI TENGAH 28 SEPTEMBER 2018 (M7.4). Pusat Penelitian dan Pengembangan Perumahan dan Permukiman Badan Penelitian dan Pengembangan Kementerian Pekerjaan Umum dan Perumahan Rakyat, 2018.
- [4] R. R. A. Rahman and A. W. Wijayanto, "PENGELOMPOKAN DATA GEMPA BUMI MENGGUNAKAN ALGORITMA DBSCAN," J. Meteorol. DAN Geofis., vol. 22, 2021.
- [5] USGS, "Earthquake Glossary." <https://earthquake.usgs.gov/learn/glossary/?termID=150> (accessed Jan. 14, 2021).
- [6] A. C. Mueller and S. Guido, Introduction to Machine Learning with Python. O'Reilly Media, Inc., 2016.
- [7] WIKIPEDIA, "k-means clustering." [https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering) (accessed Jan. 22, 2021).
- [8] D. Pelleg and A. Moore, "Accelerating Exact k-means Algorithms with Geometric Reasoning," Proc. Fifth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '99., 1999.
- [9] D. MacKay, Information Theory, Inference, and Learning Algorithms. Cambridge University Press., 2003.
- [10] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-Means Clustering Algorithm," J. R. Stat. Soc., 1979.
- [11] T. Amr, Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits. Packt Publishing, 2020.

#### BIOGRAPHY OF AUTHORS



I Made Budi Sukmadana adalah staff pengajar di Jurusan Teknik Elektro, Fak. Teknik Universitas Mataram. Ketertarikan penelitian yaitu di bidang Digital Signal Processing, Machine Learning dan Energi Baru Terbarukan.



Cipta ramadhani adalah staff pengajar di Jurusan Teknik Elektro, Fak. Teknik Universitas Mataram. Ketertarikan penelitian yaitu di bidang Digital Signal Processing dan Machine Learning