ELANG: Journal of Interdisciplinary Research
E-ISSN: 3025-2482

Analisis dan Visualisasi Data Gempa Bumi di Indonesia tahun 2008-2022 Menggunakan Google Colab

Ariel Arauna Amethyts Priyaadmaja [1], Christian Andrew [2], Muhammad Bintang Satria [3]
[1)]arielpriadmaja@gmail.com || [2)] christian.andrew0109@gmail.com || [3)] bint4ngsatria21@gmail.com

Universitas Merdeka Malang, Indonesia

Kata Kunci

Analisa Data; Google Colab; K-Means; Klasterisasi; Visualisasi Data.

****) Author Korespondensi***
christian.andrew0109@gmail.com

Abstrak

Indonesia merupakan negara yang sering mengalami gempa bumi karena letaknya berada pada pertemuan lempeng tektonik utama. Penelitian ini bertujuan untuk menganalisis distribusi spasial kejadian gempa bumi di Indonesia selama periode 2008–2022 menggunakan teknik klasterisasi. Dataset gempa bumi diperoleh dari sumber terbuka dan diolah menggunakan algoritma K-Means pada platform Google Colab. Proses klasterisasi menggunakan fitur utama seperti lintang, bujur, dan magnitudo untuk mengelompokkan aktivitas seismik ke dalam beberapa klaster. Hasil penelitian ini konsisten dengan studi sebelumnya yang menggunakan algoritma klasterisasi seperti K-Means, K-Medoids, dan BIRCH pada data seismik Indonesia, di mana zona-zona dengan aktivitas tinggi teridentifikasi di wilayah Sumatra, Jawa, Sulawesi, dan Papua. Analisis ini memberikan pemahaman lebih lanjut tentang wilayah rawan gempa dan mendukung pengembangan strategi mitigasi bahaya seismik di Indonesia.

1. Pendahuluan

Indonesia merupakan negara yang terletak pada pertemuan tiga lempeng tektonik utama dunia, yaitu Indo-Australia, Eurasia, dan Pasifik. Kondisi ini menyebabkan Indonesia memiliki tingkat aktivitas gempa bumi yang tinggi, sehingga rawan terhadap bencana gempa bumi yang dapat mengancam keselamatan jiwa dan merusak infrastruktur. Oleh karena itu, diperlukan upaya mitigasi melalui analisis data gempa bumi untuk mengidentifikasi zona rawan gempa.

Perkembangan ilmu pengetahuan dan teknologi, khususnya di bidang data mining dan machine learning, telah memungkinkan analisis data secara lebih akurat dan efisien. Algoritma klasterisasi seperti K-Means, K-Medoids, dan DBSCAN telah banyak digunakan dalam berbagai bidang untuk mengelompokkan data dengan karakteristik serupa. Septianto et al. (2025) memanfaatkan K-Means untuk klasterisasi data produksi pertanian di Kabupaten Cirebon, sedangkan studi oleh MALCOM (2025) membandingkan K-Means dan DBSCAN untuk segmentasi pelanggan transportasi Transjakarta. Prastyabudi et al. (2024) menerapkan K-Means dalam segmentasi pasar pendidikan tinggi, dan Mohmed et al. (2024) menunjukkan peran pengelolaan data pada sektor klinis.

Pada bidang mitigasi bencana, penelitian terkait klasterisasi data gempa bumi juga telah dilakukan. Mahmud (2023) menggunakan algoritma K-Means untuk klasifikasi kedalaman gempa di Sulawesi, sementara Yulian et al. (2023) memanfaatkan K-Means untuk mengelompokkan data gempa di wilayah Sulawesi. Ramadhan dan Harahap (2022) menggunakan K-Medoids untuk klasterisasi daerah rawan gempa bumi di Indonesia. Penelitian serupa juga dilakukan oleh Cahyani (2021) dengan metode hierarchical clustering dan Satriawan (2023) yang menganalisis distribusi gempa di Pulau Sumatra.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis data gempa bumi di Indonesia pada periode 2008 hingga 2022 dengan menggunakan metode K-Means clustering. Penelitian ini juga bertujuan untuk memvisualisasikan hasil klasterisasi guna mempermudah identifikasi wilayah rawan gempa secara spasial.

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan algoritma K-Means untuk mengelompokkan wilayah rawan gempa bumi di Indonesia berdasarkan data gempa bumi periode 2008–2022?
2. Bagaimana visualisasi hasil klasterisasi untuk menunjukkan zona rawan gempa di Indonesia?

Penelitian ini diharapkan dapat berkontribusi dalam mendukung mitigasi bencana gempa bumi dan perencanaan pembangunan wilayah yang lebih aman terhadap potensi bahaya seismik.

2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama yaitu:

1. Akuisisi Data
   Pada tahapan ini dilakukan pengumpulan data yang akan digunakan dalam penelitian ini. *Dataset* yang digunakan adalah data "Earthquake Occurence" yang bersumber dari website kaggle.
2. Pra Pemrosesan Data (*Pre-processing*)
   Tahapan kedua ini adalah tahapan yang dilakukan untuk menyiapkan *raw data* atau data mentah menjadi data yang siap untuk dianalisa. Pada proses ini dilakukan proses penghapusan data yang kosong, data duplikat dan penanganan pada data-data yang tidak penting.
3. Pemilihan Metode Analisis
   Metode analisis yang digunakan untuk menganalisis adalah metode klasterisasi atau *Clustering*. *Clustering* adalah metode analisis data yang dilakukan dengan mengelompokkan objek berdasarkan data yang memiliki karakteristik mirip (Satriawan, 2023). Metode klasterisasi yang digunakan pada penelitian ini adalah metode K-Means. Metode ini adalah algoritma sederhana yang digunakan pada bidang *unsupervised data mining* yang menggunakan *centroid* sebagai pusat dari klaster-klaster atau kelompok datanya (Aditya, 2024). Persamaan dari K-Means yaitu:

   $$J = \sum_{i=1}^{k} \sum_{j=1}^{n} ||x_j^{(i)} - \mu_i||^2 \quad (1)$$

   Dimana:

   J = Fungsi Objektif
   k = Jumlah Kluster
   n = Jumlah Total Titik Data
   $x_j^{(i)}$ = Titik ke-j dalam Kluster ke-i
   μ_i = Pusat Kluster ke-i
   $\|\cdot\|^2$ = Jarak Eucledian Kuadrat

   Secara umum, berikut ini adalah langkah kerja dari K-Means (Prastyabudi, 2024):
   - Menentukan nilai k sesuai jumlah kluster yang akan dibentuk
   - Menentukan *centroid* awal secara acak dari data yang ada dengan jumlah kluster sama dengan kluster awal.
   - Mencari *centroid* terdekat dari tiap data dengan menghitung jarak tiap *centroid*.
   - Mengelompokkan data yang memiliki jarak minimum dari *centroid*.
   - Mencari *centroid* baru berdasarkan rata-rata tiap kluster dari pengelompokan yang telah dilakukan.
   - Melakukan iterasi dan memulai lagi untuk menentukan *centroid* awal.
   - Pengulangan akan berhenti ketika sudah tidak ada lagi data yang bisa digeser.
4. Analisis Hasil
   Pada tahapan terakhir dilakukan analisis secara manual pada data yang telah divisualisasikan dengan cara mengamati dengan terperinci hasil visualisasi yang telah dibuat.

3. Hasil dan Pembahasan

3.1 Akuisisi Data

Pada proses ini, data yang digunakan adalah data sekunder yang memiliki format CSV (*Comma Separated Values*) berjudul "Earthquake Occurence" oleh Greg Titan yang didapatkan dari *website* kumpulan *dataset* yaitu kaggle. Data tersebut berisi sekitar 87372 gempa yang terjadi di Indonesia mulai dari tahun 2008 sampai dengan 2022. *Dataset* tersebut memiliki beberapa kolom yang berupa tanggal (*date*), jam atau waktu (*time*), lintang (*latitude*), bujur (*longitude*), kedalaman (*depth*), dan magnitudo atau kekuatan gempa (*magnitude*).

	A	B	C	D	E	F
1	date	time	latitude	longitude	depth	magnitude
2	2008-11-01	00:31:25	-0,6	98,89553	20	2,99
3	2008-11-01	01:34:29	-6,61	129,38722	30,1	5,51
4	2008-11-01	01:38:14	-3,65	127,99068	5	3,54
5	2008-11-01	02:20:05	-4,2	128,097	5	2,42
6	2008-11-01	02:32:18	-4,09	128,20047	10	2,41
7	2008-11-01	03:24:09	-3,76	127,38228	10	2,94
8	2008-11-01	03:34:47	-3,89	128,24319	10	2,22
9	2008-11-01	04:26:50	0,49	98,32848	15,9	3,85
10	2008-11-01	06:01:05	-4,26	127,66187	10	3,24

Gambar 1. *10 Baris Pertama Dataset yang ditampilkan pada Microsoft Excel.*

3.2 Pra Pemrosesan Data (*Pre-processing*)

Pada tahapan ini dilakukan proses *Data Cleaning*, proses ini berupa pembersihan dan validasi data yang penting dilakukan untuk menjamin ketepatan dan keandalan data yang telah diakuisisi, proses ini mencakup pemeriksaan dan perbaikan kesalahan pada *dataset*, ketidak konsistenan serta nilai-nilai yang hilang dari *dataset* (Mohmed, 2024).

Pada penelitian ini, proses pembersihan data dilakukan menggunakan Google Colab. Pada Google Colab digunakan *library* Python yang bernama Pandas. Dikarenakan pada *dataset* yang digunakan tidak ada data yang kosong maka proses pembersihan yang dilakukan hanya proses penghapusan data duplikat. Selain itu dilakukan juga penggabungan kolom tanggal dan waktu (*date* dan *time*) menjadi kolom tanggal dan waktu (*datetime*).

```
#Upload File
from google.colab import files
uploaded = files.upload()
```

```
Choose Files  data.csv
• data.csv(text/csv) - 4064632 bytes, last modified: 9/29/2022 - 100% done
Saving data.csv to data.csv
```

```
[ ] #Data Cleaning dan Menggabungkan Kolom Date and Time
    import pandas as pd

    df = pd.read_csv('data.csv')
    df['datetime'] = pd.to_datetime(df['date'] + ' ' + df['time'], errors='coerce')
    df = df.drop(columns=['date', 'time'])
    df = df.drop_duplicates()
    df.head()
```

	latitude	longitude	depth	magnitude	datetime
0	-0.60	98.89553	20.0	2.99	2008-11-01 00:31:25
1	-6.61	129.38722	30.1	5.51	2008-11-01 01:34:29
2	-3.65	127.99068	5.0	3.54	2008-11-01 01:38:14
3	-4.20	128.09700	5.0	2.42	2008-11-01 02:20:05
4	-4.09	128.20047	10.0	2.41	2008-11-01 02:32:18

```
[ ] #Download CSV yang sudah di clean
    df.to_csv('data_bersih.csv', index=False)
    files.download('data_bersih.csv')
```

Gambar 2. *Tampilan kode untuk upload, data cleaning dan download file.*

Karena Google Colab berbasis *browser*, maka *dataset* akan diunggah kemudian setelah itu barulah *dataset* itu bisa diolah dan dibersihkan. Selain itu pada kode juga disertakan opsi untuk mengunduh data yang telah di bersihkan.

3.3 Pemilihan Metode Analisis

Pada penelitian ini klasteriasi menggunakan K-Means masih dilakukan menggunakan Google Colab. Sama seperti sebelumnya kode yang digunakan adalah Python dan menggunakan *library* Pandas. Pada proses ini, data diklusterisasikan berdasarkan lintang (*latitude*), bujur (*longitude*) dan kekuatan gempa atau magnitudo (*magnitude*). Selain itu pada klasterisasi, data yang ada dibagi menjadi 4 klaster dengan menggunakan nilai acak 42.

```
[ ]  #Impor library yang dibutuhkan
     import pandas as pd
     from sklearn.cluster import KMeans
     import matplotlib.pyplot as plt
     import seaborn as sns

     #Membaca dan cek kolom yang akan dipakai untuk klustering (Lokasi, kedalaman dan besar/magnitudo)
     df = pd.read_csv('data_bersih.csv')

     df[['latitude', 'longitude', 'depth', 'magnitude']].describe()

     #Memilih fitur klusterisasi (Misal lokasi dan magnitudo)
     X = df[['latitude', 'longitude', 'magnitude']]

     #Metode KMeans untuk klusterisasi
     kmeans = KMeans(n_clusters=4, random_state=42) #n_cluster = jumlah kluster yang dibentuk; random state = nilai acak
     df['cluster'] = kmeans.fit_predict(X)
```

Gambar 3. *Tampilan kode untuk K-means.*

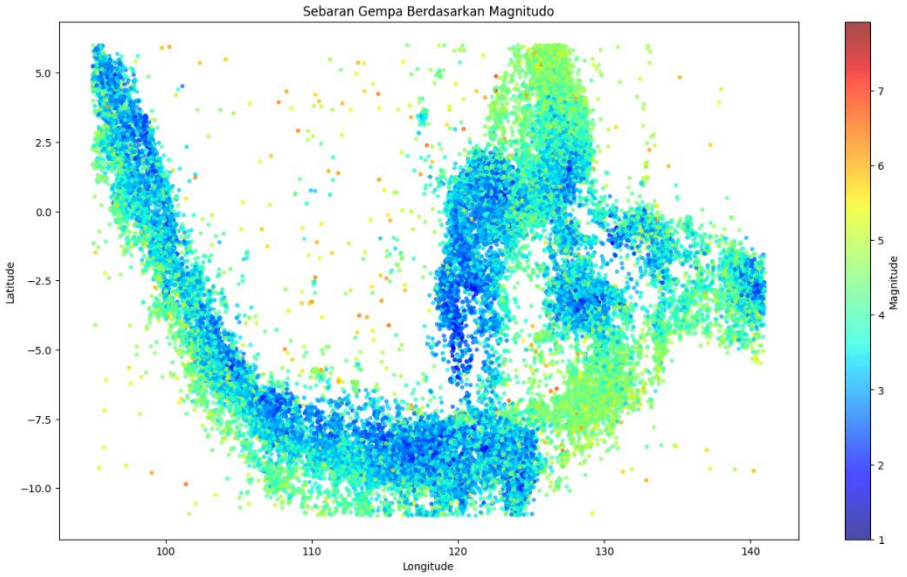
Setelah di klasterisasi, data yang ada divisualisasikan menjadi 2 grafik untuk memudahkan proses analisis. Pada 2 grafik tersebut bagian sumbu x letak bujur atau *longitude*, sedangkan sumbu *y* menunjukkan letak lintang atau *latitude*. Karena itu letak titik nantinya akan menunjukkan lokasi dari gempa yang ada. Untuk hasil visualisasinya titik-titik pada grafik 1 akan diberi warna berdasarkan magnitudo atau kekuatan gempanya. Sedangkan untuk grafik 2, titik-titik akan diwarna berdasarkan kluster lokasinya.

```
#Visualisasi
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(16, 9)) #Ukuran tampilan
scatter = plt.scatter(
    df['longitude'],
    df['latitude'],
    c=df['magnitude'],      #Warna berdasarkan magnitudo
    cmap='jet',             #Skema warna
    s=10,                   #Ukuran titik
    alpha=0.7               #Transparansi
)
plt.colorbar(scatter, label='Magnitude')  # Menampilkan skala warna
plt.title("Sebaran Gempa Berdasarkan Magnitudo")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.show()

#Kluster berdasarkan lokasi
plt.figure(figsize=(16, 9))
sns.scatterplot(data=df, x='longitude', y='latitude', hue='cluster', palette='Set2')
plt.title("Cluster Gempa Berdasarkan Lokasi")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.legend(title='Cluster')
plt.show()
```

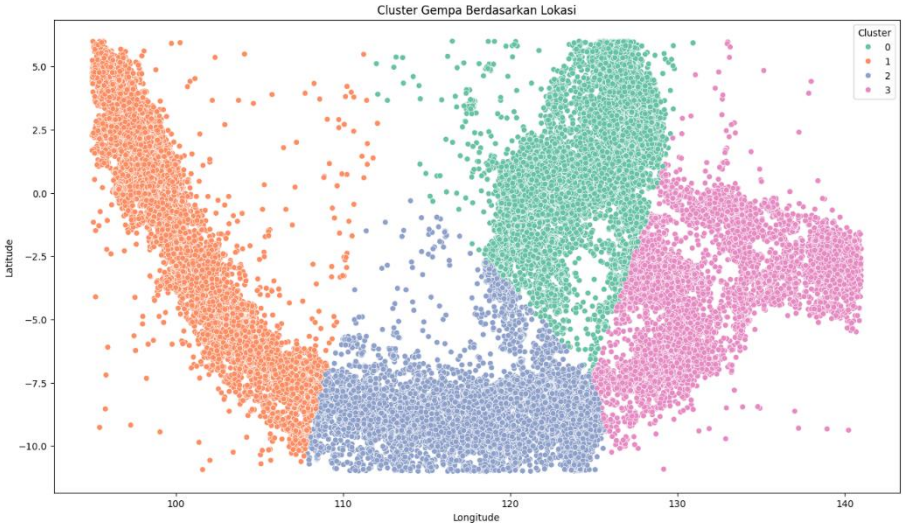
Gambar 4. *Tampilan kode untuk visualisasi grafik 1 dan 2.*

3.4 Analisis Hasil

Dari proses yang sebelumnya, berikut ini adalah hasil visualisasi dari *dataset* yang digunakan:



Gambar 5. *Grafik 1.*



Gambar 6. *Grafik 2.*

Dari visualisasi dapat dilihat jika memang dilihat berdasarkan lokasi titik-titiknya memang menunjukkan bentuk dari pulau-pulau di Indonesia. Berikut ini gambar peta Indonesia sebagai perbandingan:

Gambar 7. *Peta Indonesia dari BPK Provinsi Bangka Belitung.*

Dengan membandingkan kedua grafik dan peta Indonesia, jika dikelompokkan berdasarkan 4 kluster dapat dilihat bahwa gempa di Indonesia dibagi menjadi kluster 1 yang berisi Pulau Sumatera dan Pulau Jawa bagian Barat, kluster 2 yang berisi Pulau Jawa hingga ke Nusa Tenggara Timur, kluster 3 yaitu Pulau Sulawesi dan Maluku Utara, dan kluster 4 yaitu Papua dan Maluku. Selain itu juga bisa dilihat jika kebanyakan gempa yang terjadi berada di daratan ataupun di sekitar pulau-pulau di Indonesia kecuali Pulau Kalimantan yang jika dilihat antara peta dan grafiknya tidak memiliki titik-titik gempa.

Sedangkan untuk magnitudo gempa di Indonesia rata-rata memiliki magnitudo 1 hingga 5. Dimana gempa berskala magnitudo 5 berada pada Utara Pulau Sumatra, Utara Pulau Jawa, Nusa Tenggara dan Sulawesi. Sedangkan untuk magnitudo 1 berada pada Barat Pulau Sumatra, Selatan Pulau Jawa dan pada Maluku. Meskipun hanya sedikit pada grafik juga menunjukkan ada beberapa lokasi yang memiliki magnitudo gempa lebih dari 5 yang kebanyakan ada di daerah perairan atau pada letak pulau Kalimantan.

4 Kesimpulan

4.1 Simpulan

Berdasarkan hasil analisis dan visualisasi data gempa bumi di Indonesia periode 2008 hingga 2022, dapat disimpulkan bahwa metode K-Means clustering berhasil mengelompokkan wilayah rawan gempa menjadi empat klaster utama. Klaster-klaster tersebut menunjukkan persebaran gempa bumi yang dominan di wilayah Sumatra, Jawa, Sulawesi, Maluku, dan Papua. Hasil visualisasi menunjukkan bahwa sebagian besar gempa bumi terjadi di sekitar zona subduksi dan jalur pegunungan tektonik aktif. Selain itu, analisis magnitudo menunjukkan bahwa gempa bumi di Indonesia umumnya memiliki kekuatan antara 1 hingga 5 skala Richter, dengan beberapa kejadian yang mencapai lebih dari 5 magnitudo.

4.2 Saran

Penelitian selanjutnya disarankan untuk:

1 Menggunakan algoritma klasterisasi lain seperti DBSCAN atau Hierarchical Clustering untuk membandingkan hasil segmentasi wilayah rawan gempa.
2 Mengintegrasikan data kedalaman gempa dan parameter geologis lainnya untuk meningkatkan akurasi hasil klasterisasi.
3 Menerapkan metode prediksi gempa berbasis machine learning guna mendukung mitigasi bencana secara real-time.
4 Memperbaharui dataset secara berkala agar analisis data gempa tetap relevan dengan kondisi terbaru.

5 Referensi

Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). Klasterisasi data produksi pertanian di Kabupaten Cirebon dengan algoritma K-Means. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2). https://doi.org/10.23960/jitet.v13i2.6174

Perbandingan algoritma DBSCAN dan K-Means dalam segmentasi pelanggan pengguna transportasi publik Transjakarta menggunakan metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*. Diakses 3 Juli 2025, dari https://journal.irpi.or.id/index.php/malcom/article/view/1516

Prastyabudi, W. A., Alifah, A. N., & Nurdin, A. (2024). Segmenting the higher education market: An analysis of admissions data using K-Means clustering. *Procedia Computer Science*, 234, 96–105. https://doi.org/10.1016/j.procs.2024.02.156

Mohmed, S., Makutam, V., Priya, M., & Javid, A. (2024). *An overview on clinical data management and role of Pharm.D in clinical data management*. ResearchGate. Diakses 3 Juli 2025, dari https://www.researchgate.net/publication/383665730_AN_OVERVIEW_ON_CLINICAL_DATA_MANAGEMENT_AND_ROLE_OF_PHARMD_IN_CLINICAL_DATA_MANAGEMENT

BPK Perwakilan Provinsi Bangka Belitung. (n.d.). *Gambar peta Indonesia*. Diakses 3 Juli 2025, dari https://babel.bpk.go.id/gambar-peta-indonesia-1/

Mahmud, M. S. (2023). Klasifikasi kedalaman kejadian gempa menggunakan algoritma K-Means Clustering: Studi kasus kejadian gempa di Sulawesi. *Jurnal Fisika dan Teknologi*, 5(2).

Yulian, A., Yuwono, R., & Wibowo, D. (2023). Pemanfaatan algoritma K-Means dalam klasterisasi gempa Sulawesi. *Faktor Exacta*, 14(1), 45–52.

Ramadhan, R. A., & Harahap, A. (2022). Klasterisasi daerah rawan gempa bumi di Indonesia menggunakan algoritma K-Medoids. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2).

Cahyani, F. (2021). *Analisis kelompok data gempa bumi di Indonesia menggunakan algoritma K-Means dan Agglomerative Hierarchical K-Means* (Skripsi, Institut Teknologi Sepuluh Nopember).

Satriawan, H. P. (2023). Analisis klaster distribusi gempa Pulau Sumatra. *Jurnal Desimal*, 4(2).

Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). Klasterisasi data produksi pertanian di Kabupaten Cirebon dengan algoritma K-Means. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2), Artikel No. 2. https://doi.org/10.23960/jitet.v13i2.6174

Aditiya, S. & Raka, Y. (2024) Perbandingan algoritma DBSCAN dan K-Means dalam segmentasi pelanggan pengguna transportasi publik Transjakarta menggunakan metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*. Diakses 3 Juli 2025, dari https://journal.irpi.or.id/index.php/malcom/article/view/1516

Prastyabudi, W. A., Alifah, A. N., & Nurdin, A. (2024). Segmenting the higher education market: An analysis of admissions data using K-Means clustering. *Procedia Computer Science*, 234, 96–105. https://doi.org/10.1016/j.procs.2024.02.156

Mohmed, S., Makutam, V., Priya, M., & Javid, A. (2024). *An overview on clinical data management and role of Pharm.D in clinical data management* [PDF]. ResearchGate. Diakses 3 Juli 2025, dari https://www.researchgate.net/publication/383665730_AN_OVERVIEW_ON_CLINICAL_DATA_MANAGEMENT_AND_ROLE_OF_PHARMD_IN_CLINICAL_DATA_MANAGEMENT

BPK Perwakilan Provinsi Bangka Belitung. (n.d.). *Gambar peta Indonesia*. Diakses 3 Juli 2025, dari https://babel.bpk.go.id/gambar-peta-indonesia-1/