

BAB II

TINJAUAN PUSTAKA

2.1 GEMPA BUMI

Gempa bumi adalah getaran atau guncangan pada permukaan bumi yang disebabkan oleh pelepasan energi yang terjadi di dalam bumi. Penyebab terjadinya gempa bumi dapat bervariasi, termasuk aktivitas lempeng tektonik, letusan gunung berapi, pelepasan gas alam atau air tanah di dalam bumi, dan aktivitas manusia seperti pengeboran sumur minyak atau gas (Nias Raya et al., 2023). Gempa bumi merupakan suatu peristiwa alam yang disebabkan oleh keluarnya energi secara tiba-tiba dari dalam kerak bumi, peristiwa ini menimbulkan getaran dan guncangan pada bumi. Beberapa penyebab terjadinya gempa bumi adalah aktivitas lempeng tektonik, pergerakan lempeng tektonik, dan deformasi kerak bumi (Tantyoko et al., 2023).

Berdasarkan karakteristiknya ada 3 macam gempa bumi (Sungkawa, 2016):

1. Gempa bumi runtuh
Gempa bumi yang disebabkan oleh runtuhnya lubang-lubang interior bumi, misalnya runtuhnya lorong tambang dan lorong sebuah gua. Gempa bumi ini merupakan yang paling kecil getarannya
2. Gempa bumi vulkanik
Gempa bumi akibat aktivitas gunung api, yaitu akibat gerakan magma dari dalam bumi (batholit) yang naik ke atas (lubang kepundan). Getaran gempa ini relatif kecil
3. Gempa bumi tektonik
Gempa yang terjadi karena pelepasan tenaga akibat pergeseran sesar atau kekenyalan elastis pada daerah tumbukan lempeng samudra dengan lempeng benua.

Gempa bumi dapat dikategorikan menjadi tiga jenis berdasarkan letak titik di dalam bumi (hiposentrumnya) (Nias Raya et al., 2023) :

1. Gempa bumi dalam, terjadi ketika hiposentrum berada pada kedalaman >300 km di bawah permukaan bumi.

2. Gempa bumi menengah, terjadi ketika hiposentrum berada pada kedalaman antara 70 km hingga 300 km di bawah permukaan bumi.
3. Gempa bumi dangkal, terjadi ketika hiposentrum berada pada kedalaman <70 km di bawah permukaan bumi.

Ukuran gempa dapat dinyatakan dalam skala Richter (M) atau skala Modified Mercalli (MMI). Skala Richter mengukur Magnitude gempa berdasarkan amplitudo yang terjadi sehingga lebih objektif (Nias Raya et al., 2023).

Tabel 2. 1 Deskripsi Ukuran Skala *Richter*

Skala <i>Richter</i>	Efek Gempa
< 2,0	Tidak terasa, gempa kecil
2,0 – 2,9	Tidak terasa, terekam alat
3,0 – 3,9	Seringkali terasa, tetapi jarang kerusakan
4,0 – 4,9	Dapat diketahui getarannya, kerusakan tidak signifikan
5,0 – 5,9	Menyebabkan kerusakan pada area kecil
6,0 – 6,9	Dapat merusak area hingga jarak sekitar 160 Km
7,0 – 7,9	Menyebabkan kerusakan pada area luas

Skala *Modified Mercalli* mengukur Intensitas gempa berdasarkan efeknya terhadap manusia atau bangunan sehingga lebih bersifat subjektif (Nias Raya et al., 2023). Jika dibandingkan antara skala Richter dengan skala Modified Mercalli diperoleh hubungan :

Tabel 2. 2 Deskripsi Ukuran Skala *Modified Mercalli Intensity* (MMI)

Skala <i>Richter</i>	Intensitas (MMI)	Pengaruh-pengaruh Tipikal
≤ 2	I – II	Pada umumnya tidak terasa.
3	III	Terasa di dalam rumah, tidak ada kerusakan

4	IV – V	Terasa oleh banyak orang, barang-barang bergerak, Tidak ada kerusakan struktural
5	VI – VII	Terjadi beberapa kerusakan struktural, seperti Retakretak pada dinding.
6	VII – VIII	Kerusakan menengah, seperti hancurnya dinding
7	IX – X	Kerusakan besar, seperti runtuhnya bangunan.
≥ 8	XI – XII	Rusak total atau hampir hancur total.

2.2 DATA MINING

Data mining merupakan salah satu cara untuk mendapatkan informasi yang tersimpan pada databased yang berjumlah besar (Nasution et al., 2021). Proses *Data mining* terdiri dari pengumpulan data, ekstraksi data, analisis data dan statistik data. Dan biasa dikenal dengan *Knowledge Discovery Of Databases (KDD)* dimulai dari tahap *selecting, preprocessing, transformation, data mining, evaluation/interpretation* (Widaningsih, 2019). Penggunaan *data mining* untuk menghitung probabilitas data bencana adalah pendekatan yang dapat membantu dalam mengidentifikasi pola-pola dan hubungan antara variabel-variabel yang berkontribusi terhadap terjadinya bencana alam. (Althaf & Rianto, 2023).

Berikut ini merupakan proses *KDD* pada data mining (Alfiqra & Yogi, 2018):

1. *Data Selection*, Proses pemilihan data yang dilakukan adalah dengan memilih data penelitian yang akan diproses dalam proses *data mining*.
2. *Preprocessing/Cleaning*, Setelah data dipilih, kemudian proses pembersihan data untuk menghapus data yang tidak berhubungan dalam penelitian. Duplikasi data dihapus, data yang tidak konsisten diperiksa, dan kesalahan data, seperti kesalahan ketik, dikoreksi sebagai bagian dari metode prapemrosesan.
3. *Data Transformation*, Tahap ini merupakan tahap untuk mengubah bentuk data, pada penelitian ini, penulis menggunakan teknik *aggregation*, atau proses peringkasan pada data mentah.

4. *Data Mining*, Tahap ini merupakan proses pengolahan data berdasarkan algoritma sesuai dengan teknik data mining sesuai dengan kebutuhan dan fungsinya nantinya.
5. *Evaluation/Interpretation*, Tahap terakhir yaitu evaluasi atau proses menginterpretasikan hasil rule yang didapatkan dari Teknik data mining yang berguna untuk menarik kesimpulan dari suatu analisa atau proses data mining tersebut.

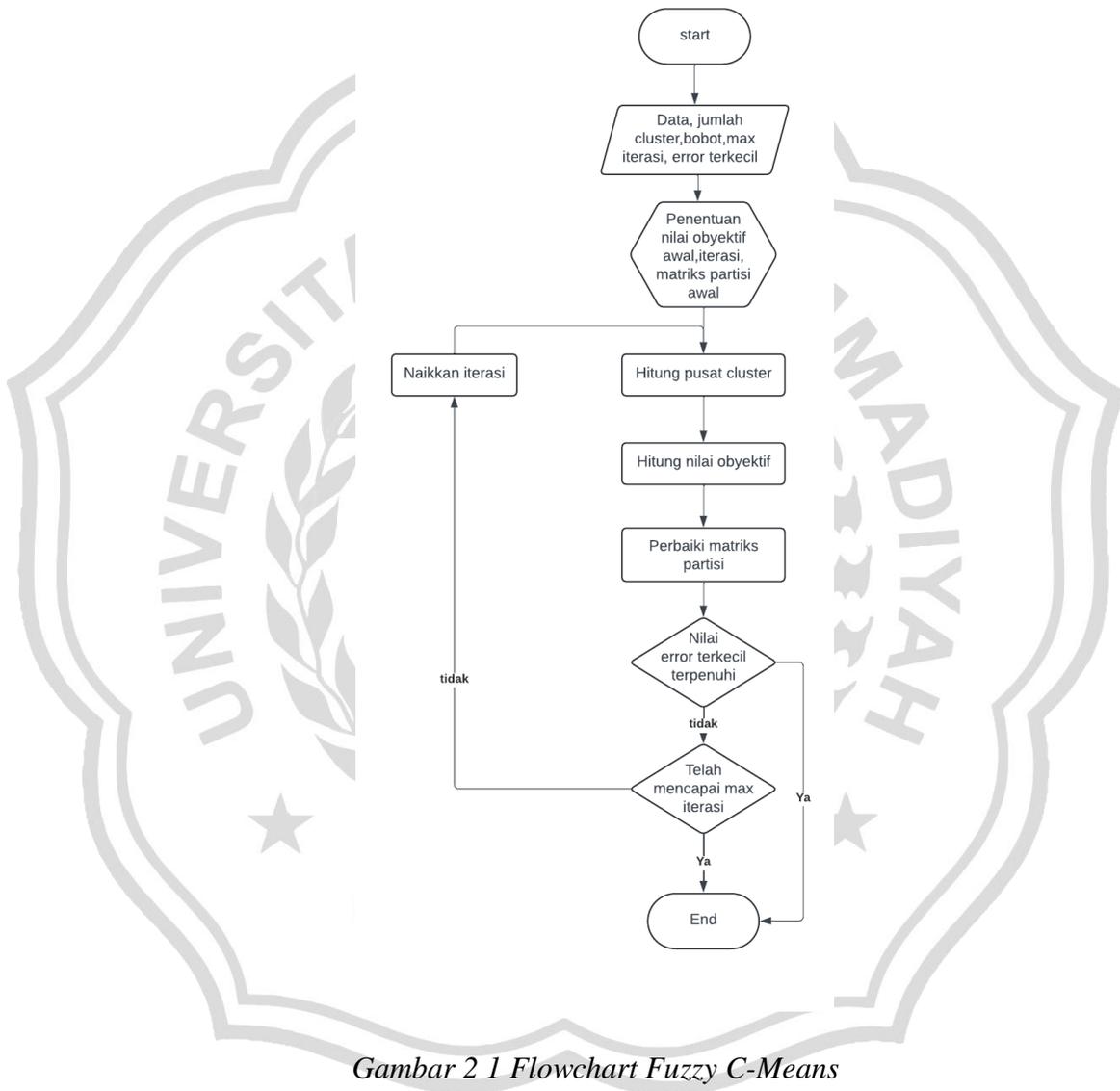
2.3 CLUSTERING

Clustering adalah metode yang digunakan dalam mendata cara kerjanya mencari dan mengelompokkan data yang memiliki kemiripan dan karakteristik data antara satu dengan yang lain dengan datanya (Yuni Radana Sembiring et al., 2021). Tujuan *clustering* untuk menentukan kelompok dalam sekumpulan data yang tidak berlabel, karena clustering merupakan suatu metode *unsupervised* dan tidak terdapat suatu kondisi awal untuk sejumlah *cluster* yang mungkin terbentuk dalam sekumpulan data, maka dibutuhkan suatu evaluasi hasil *clustering* (Paembonan et al., 2021).

2.4 ALGORITMA FUZZY C-MEANS

Fuzzy C-Means didasarkan pada teori logika *fuzzy*. Lotfi Zadeh (1965) memperkenalkan teori dari pemodelan ini pertama kali dimana keanggotaan dari data tidak secara tegas diberi nilai dengan 0 (tidak menjadi anggota cluster) dan nilai 1 (menjadi anggota cluster), namun dengan sebuah nilai derajat keanggotaan yang batasan nilainya 0 sampai 1. Tahapan awal dari konsep dasar *Fuzzy C-Means* yang paling awal adalah dengan menentukan pusat *cluster* (*centroid*) yang akan mengidentifikasi lokasi/ruang rata-rata untuk tiap-tiap *cluster*. Dalam kondisi awal, pusat cluster ini belum dapat dikatakan akurat hal ini diakibatkan oleh setiap data memiliki derajat keanggotaan untuk masing-masing cluster. Perbaikan terhadap pusat *cluster* (*centroid*) dan masing-masing nilai keanggotaan data dengan perulangan, akan terlihat bahwa pusat *cluster* (*centroid*) akan bergerak mendekati ruang/lokasi yang tepat (Abdulrohman et al., 2023).

Flowchart Fuzzy C-Means dapat dilihat pada Gambar 2.1, dapat dilihat proses-proses yang dilakukan pada algoritma *FCM* mulai dari menentukan variabel perhitungan, menghitung pusat kluster, menghitung nilai obyektif, memperbaiki matriks partisi sampai penentuan kondisi berhenti (Astuti & Nurwiyati, 2020).



Gambar 2 1 Flowchart Fuzzy C-Means

Langkah-langkah metode *Fuzzy C-Means* Sebagai berikut

1. Input data

Langkah ini dapat dilakukan dengan mengasumsikan sejumlah data dalam set X , berupa matriks berukuran $n \times m$ ($n =$ jumlah data, $m =$ jumlah atribut). x_{ij} adalah data sampel ke- i ($i = 1, 2, 3, \dots, n$) dan atribut ke- j ($j = 1, 2, 3, \dots, m$).

2. Tentukan jumlah *cluster*

Tentukan jumlah *cluster* (K), pangkat untuk matriks partisi (w), maksimum iterasi ($MaxIter$), *error* terkecil yang diharapkan (ξ), fungsi objektif awal ($P_0=0$), dan iterasi awal ($t=1$).

3. Bangkitkan nilai random

Bangkitkan matriks *fuzzy pseudo-partition* yang diinisialisasikan dengan memberikan nilai sembarang dalam jangkauan $[0,1]$ dan jumlah untuk setiap data (baris) adalah 1. U_{ik} adalah matriks data ke- i ($i = 1,2,3,\dots,n$) dan cluster ke- k ($k=1,2,\dots,k$).

4. Bangkitkan bilangan random sebagai elemen – elemen matriks partisi awal U , dengan ketentuan

$$U = \begin{bmatrix} u_{1,1} & \dots & u_{1,k} \\ \vdots & \ddots & \vdots \\ u_{n,1} & \dots & u_{n,k} \end{bmatrix} \quad (2.1)$$

Matriks partisi pada fuzzy harus memenuhi kondisi persamaan sebagai berikut:

$$u_{ij} \in [0,1]; (1 \leq i \leq n, 1 \leq j \leq k), \quad (2.2)$$

$$\sum_{j=1}^k u_{ij} = 1; (1 \leq i \leq n), \quad (2.3)$$

$$0 < \sum_{i=1}^n u_{ij} < \infty; (1 \leq j \leq k) \quad (2.4)$$

5. Hitung pusat cluster ke- k menggunakan persamaan (2.5) :

$$V_{kj} = \frac{\sum_{i=1}^n (u_{ij})^w}{\sum_{i=1}^n (u_{ij})^w} \quad (2.5)$$

Hasil dari perhitungan ini berupa matriks (2.6) :

$$V = \begin{bmatrix} v_{1,1} & \dots & v_{1,k} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \dots & v_{n,k} \end{bmatrix} \quad (2.6)$$

Dengan :

x_{kj} : Komponen ke-j dari pusat cluster ke-i.

μ_{ki} : Derajat keanggotaan dari data ke-k terhadap cluster ke-i.

x_{kj} : Komponen ke-j dari data ke-k.

6. Hitung Fungsi objektif

Hitung fungsi objektif pada iterasi ke-t menggunakan persamaan (2.7):

$$P_t = \sum_{k=1}^n \sum_{i=1}^c ([\sum_{j=1}^m (x_{kj} - x_{ij})^2] (\mu_{ki})^m) \quad (2.7)$$

Keterangan :

P_t : fungsi objektif pada iterasi ke-t

m : Parameter fuzziness yang sama seperti m dalam notasi sebelumnya ($m > 1$).

n : Jumlah data.

c : Jumlah cluster.

7. Hitung perubahan matriks partisi

$$\mu_{ki} = \frac{[\sum_{j=1}^m (x_{kj} - x_{ij})^2]^{-1}}{\sum_{i=1}^c [\sum_{j=1}^m (x_{kj} - x_{ij})^2]^{-1}} \quad (2.8)$$

Keterangan:

n : dalam perhitungan ini n berarti Jumlah komponen dalam data.

m : Parameter fuzziness (biasanya

Bila dilakukan perhitungan tersebut bisa dilakukan dengan cara menghitung jarak *Euclidean* terlebih dahulu :

$$d_{ki} = \sqrt{\sum_{j=1}^m (x_{kj} - x_{ij})^2} \quad (2.9)$$

Dengan :

d_{ki} : adalah jarak Euclidean antara data kkk dan centroid cluster i

Sehingga dapat merujuk pada rumus dasar yang di gunakan pada persamaan 2.10 yaitu :

$$\mu_{ki} = \frac{1}{\sum_{i=1}^c \frac{d_{ki}^{-2}}{d_{ki}^{-2}}} \quad (2.10)$$

$\Sigma_{i=1}^n (x_i)$



8. Cek kondisi berhenti

Iterasi akan berhenti perubahan nilai fungsi objektif kurang dari ambang batas yang telah ditentukan atau iterasi telah melebihi maksimal iterasi berikut persamaan (2.11):

$$(|P_{k+1} - P_k - 1| < \epsilon) \quad (k > k_{max}) \quad (2.11)$$

2.5 SILHOUETTE COEFICIENT

Merupakan metode yang digunakan untuk kluster pada tiap objek. *Silhouette Coeficient* dikembangkan pertama kali oleh Rosseeuw dan Kaufman. Metode ini memberikan gambaran singkat tentang seberapa baik objek pada klasternya (Ulinnuha et al., 2020). *Silhouette score* adalah sebuah metrik evaluasi yang digunakan untuk mengukur sejauh mana setiap objek dalam sebuah kelompok atau kluster saling berdekatan dan terpisah dari kelompok atau kluster lainnya. Metrik ini memberikan informasi tentang kualitas pembagian kluster, dimana nilai siluet yang lebih tinggi menunjukkan bahwa objek dalam suatu kluster saling berdekatan dengan baik dan terpisah dengan jelas dari kluster lainnya (Rizuan et al., 2023).

Tahapan perhitungan *silhouette coefficient* sebagai berikut :

1. Menghitung rata-rata jarak suatu data dengan data lain dalam suatu cluster yang sama menggunakan persamaan (2.12)

$$s(i) = \frac{1}{|A_i| - 1} + \sum_{A_j \neq A_i} d(i, j) \quad (2.12)$$

Keterangan :

$d(i, j)$: Perbedaan rata-rata pada data (i) ke semua data lain di cluster A

$d(i, j)$: Jarak antara data i dan data j

2. Menghitung rata-rata jarak data tersebut dengan semua data di *cluster* lain menggunakan persamaan (2.13) :

$$s(i, j) = \frac{1}{|A_j|} + \sum_{k \in A_j} d(i, k) \quad (2.13)$$

Keterangan :

$\bar{x}(i)$: Perbedaan rata-rata pada data (i) ke seluruh data

\bar{x} : Cluster lain selain A

- Memilih nilai jarak dengan nilai paling kecil atau minimum menggunakan persamaan (2.14)

$$\bar{x}(i) = \frac{\sum_{j \neq A} d(i, j)}{n - 1} \quad (2.14)$$

- Hitung nilai silhouette coefficient dengan persamaan (2.15):

$$s(i) = \frac{\bar{x}(i) - \bar{x}(A)}{\max\{\bar{x}(i), \bar{x}(A)\}} \quad (2.15)$$

Keterangan :

$s(i)$: Nilai *Silhouette Coefficient*

$\bar{x}(i)$: Perbedaan rata-rata pada data (i) ke semua data pada cluster berbeda

$\bar{x}(A)$: Perbedaan rata-rata pada data (i) ke semua data lain di cluster A

Range nilai s(i) yakni antara -1 dan 1, interpretasi nilai tersebut yaitu:

S(i) = -1, data ke- i digolongkan lemah (dekat pada cluster B daripada A).

S(i) = 0, diartikan data ke- i berada di tengah dua kluster (A dan B)

S(i) = 1, diartikan data ke- i digolongkan baik

Interpretasi nilai *Silhouette Coefficient* di tunjukkan pada tabel 2.3

Tabel 2. 3 Interpretasi nilai Silhouette Coefficient

<i>Silhouette Coefficient</i>	Interpretasi
≤ 0.25	Tidak terstruktur
0.26-0.50	Hasil struktur lemah
0.51-0.70	Hasil struktur baik
0.71-1.00	Hasil struktur kuat

2.6 PENELITIAN TERKAIT

Penelitian terkait yang menjadi landasan teori dalam penelitian ini membahas metode dan teknik yang serupa dengan permasalahan yang dihadapi. Daftar penelitian terkait ditunjukkan pada tabel 2.4:

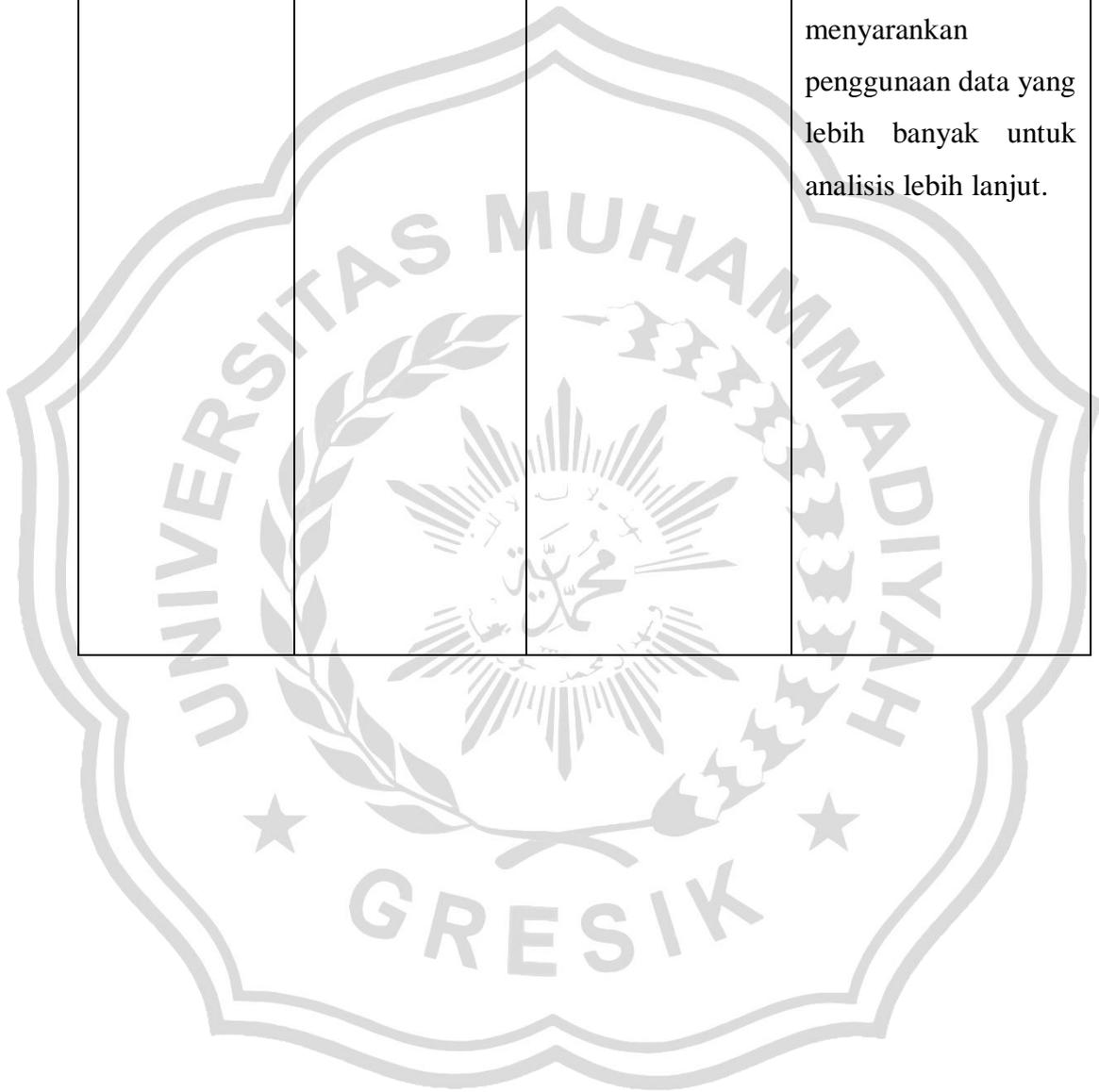
Tabel 2. 4 Penelitian Terkait

Landasan Literatur	Metode Yang Digunakan	Masalah	Hasil Penelitian
"Implementasi Algoritma Fuzzy C-Means Pada Pengelompokan Siswa Pkl". Abdulrohim, U., Adawiah, S., dan Gunawan, H.	Metode yang digunakan adalah metode Fuzzy C-Means	Bagaimana mengelompokkan siswa untuk penempatan PKL di perusahaan yang sudah ditentukan oleh pihak kepala sekolah	Hasil penelitian ini menunjukkan bahwa SMK Mutiara Salawu berhasil mengimplementasikan sistem pengelompokan siswa untuk Praktik Kerja Lapangan (PKL) menggunakan algoritma Fuzzy C-Means. Sistem ini dirancang untuk mengelompokkan siswa berdasarkan nilai rapor mereka, sehingga penempatan PKL dapat lebih sesuai dengan kemampuan dan kompetensi siswa. Dengan menggunakan algoritma ini, sekolah

		<p>dapat melakukan pengelompokan data secara efektif, yang memungkinkan penempatan siswa di lokasi PKL yang lebih tepat. Penelitian ini menekankan pentingnya pendekatan yang lebih sistematis dalam pengelompokan siswa, mengingat sebelumnya sekolah tidak memiliki sistem khusus untuk tujuan ini. Implementasi sistem ini diharapkan dapat meningkatkan proses penempatan siswa secara keseluruhan.</p>
--	--	---

<p>"Penerapan <i>Market Basket Analysis</i> Menggunakan Proses Kdd (Knowledge Discovery In Database) Sebagai Strategi Penjualan Produk Swalayan (Studi Kasus : Swalayan X)". Alfiqra, F., dan Yogi, A. (2018).</p>	<p>Metode yang digunakan dalam penelitian ini adalah Market Basket Analysis dengan algoritma <i>FP-Growth</i>. Proses KDD (Knowledge Discovery in Database) juga digunakan dalam penelitian ini.</p>	<p>Menganalisis keranjang belanja konsumen untuk menentukan strategi pemasaran dengan cara mengetahui produk yang di beli secara bersamaan.</p>	<p>Hasil penelitian ini menunjukkan bahwa dari analisis 100 transaksi belanja di sebuah swalayan di Yogyakarta menggunakan metode <i>Market Basket Analysis</i> dengan algoritma <i>FP-Growth</i>, ditemukan 10 aturan asosiasi. Aturan-aturan ini memiliki nilai <i>support</i> tertinggi sebesar 12% dan <i>confidence</i> tertinggi sebesar 66,7%, yang menunjukkan adanya hubungan kuat antara produk snack dan minuman. Berdasarkan hasil analisis ini, dua strategi penjualan diusulkan: pertama, mengatur tata letak produk dengan mendekatkan barang yang sering dibeli bersamaan; kedua, membuat katalog</p>
--	--	---	---

promosi yang menawarkan diskon pada produk yang jarang dibeli. Penelitian ini juga menyarankan penggunaan data yang lebih banyak untuk analisis lebih lanjut.



<p><i>"Analisis Probabilitas Bencana Alam dengan Penerapan Data Mining Menggunakan K-Means dan Linier Regression"</i>. Althaf, M. A. P. P., dan Rianto. (2023).</p>	<p>Metode yang digunakan dalam penelitian ini adalah <i>Data mining</i> dengan teknik <i>K-Means</i> untuk dan linear regression</p>	<p>Menganalisis data dari kejadian bencana alam di wilayah Indonesia menggunakan metode data mining, dengan teknik k-means untuk klasterisasi dan <i>linear regression</i> untuk prediksi data bencana alam</p>	<p>Hasil penelitian menunjukkan bahwa setelah melakukan klasterisasi data menggunakan metode k-means, didapatkan tiga klaster wilayah berdasarkan jenis bencana yang sering terjadi: klaster banjir (Jawa, Sulawesi, Bali, dan Nusa Tenggara), klaster tanah longsor (Sumatera, Kepulauan Riau, Papua, dan Maluku), dan klaster puting beliung (Kalimantan). Probabilitas kejadian bencana di wilayah-wilayah ini dihitung menggunakan model linear regression. Penelitian ini juga menemukan bahwa wilayah Jawa memiliki probabilitas peningkatan bencana banjir yang tinggi, sedangkan wilayah Kepulauan Riau</p>
---	--	---	---

memiliki probabilitas peningkatan tertinggi untuk bencana tanah longsor dan puting beliung. Namun, hasil ini tidak memberikan efek signifikan pada kejadian aktualnya karena data yang digunakan hanya berdasarkan jumlah kejadian.

<p>"Visualisasi Area Antar Pedukuhan Hasil Clustering Penduduk Miskin Menggunakan Fuzzy C-Means". Astuti, F. D., dan Nurwiyati, F. W. (2020).</p>	<p>Metode yang digunakan adalah <i>Fuzzy C-Means</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah kemiskinan di Kecamatan Bantul, Kabupaten Bantul, dan kesulitan dalam pendistribusian bantuan bagi warga miskin karena penentuan penduduk miskin didasarkan pada banyak aspek yang selama ini hanya disajikan dalam bentuk teks dan belum divisualisasikan. Hal ini menyebabkan bantuan terkadang tidak tepat sasaran.</p>	<p>Hasil penelitian ini menunjukkan bahwa visualisasi hasil clustering berdasarkan area antar pedukuhan pada data penduduk miskin di Kecamatan Bantul dapat membantu dalam melihat penduduk yang kekurangan aspek tertentu di setiap pedukuhan. Penelitian ini menggunakan metode <i>Fuzzy C-Means</i> untuk melakukan clustering. Visualisasi ini memungkinkan untuk menampilkan status penduduk miskin per keluarga dan membantu dalam mengidentifikasi penduduk yang paling banyak kekurangan aspek tertentu. Dengan demikian, hasil clustering ini dapat mempermudah</p>
---	---	---	--

dalam pendistribusian bantuan agar lebih tepat sasaran.



<p><i>"Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar". Nasution, R. W., Kirana, I. O., Gunawan, I., & Sari, I. P. (2021).</i></p>	<p>Metode yang digunakan adalah Data mining dengan algoritma K-Means.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah pengelompokan kembali data minat konsumen terhadap jasa pengiriman di kantor JNE Pematangsiantar. Tujuannya adalah untuk menarik kembali minat masyarakat dalam menggunakan jasa pengiriman JNE tersebut, sehingga dapat meningkatkan keinginan masyarakat untuk selalu menggunakan jasa tersebut.</p>	<p>Hasil penelitian ini menunjukkan bahwa data mining diterapkan untuk mengelompokkan minat konsumen terhadap pengguna jasa pengiriman di PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar. Data yang digunakan adalah data dari tahun 2019 selama tiga bulan, yaitu September, Oktober, dan November, dengan cabang sebanyak 27 kota/kabupaten/kecamatan. Pengumpulan data dilakukan dengan mengambil data yang telah ada di PT. JNE dan melakukan wawancara dengan pihak PT. JNE Pematangsiantar. Penelitian ini menggunakan dua cluster lokasi dan dua</p>
--	---	---	---

			<p>iterasi. Setelah hasil dari setiap <i>cluster</i> didapatkan, pusat <i>cluster</i> baru dihitung berdasarkan data member tiap <i>cluster</i>. Hasil dari iterasi kedua sama dengan iterasi pertama, yang menunjukkan bahwa pengelompokan telah stabil.</p> <p>Tujuan dari pengelompokan ini adalah untuk menarik kembali minat masyarakat dalam menggunakan jasa pengiriman JNE, sehingga dapat meningkatkan keinginan masyarakat untuk selalu menggunakan jasa tersebut.</p>
--	--	--	--

<p>"Klasifikasi Data Gempa Bumi di Pulau Sumatera Menggunakan Algoritma Naïve Bayes". Nias Raya, U., Duha, T., Laia, M., Khorul Huda, A., dan Jasuma, A. (2023).</p>	<p>Metode yang digunakan adalah algoritma Naïve Bayes</p>	<p>Masalah yang diangkat dalam penelitian ini adalah klasifikasi data gempa bumi di Pulau Sumatera menggunakan algoritma Naïve Bayes, dengan fokus pada pengelompokan gempa menjadi tiga kategori: gempa dangkal, sedang, dan dalam.</p>	<p>Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk mengklasifikasikan data gempa bumi di Pulau Sumatera dengan baik. Data gempa bumi diklasifikasikan berdasarkan kedalaman hiposentrum menjadi tiga kategori: gempa dangkal, gempa menengah, dan gempa dalam. Proses preprocessing data dilakukan dengan normalisasi fitur untuk menghindari bias akibat perbedaan skala. Hasil klasifikasi menunjukkan akurasi yang baik, dengan nilai <i>AUC</i> 0.962, <i>CA</i> 0.936, dan <i>F1-Score</i> 0.939. Evaluasi dilakukan menggunakan</p>
--	---	--	--



			<p><i>Confusion Matrix</i> untuk mengukur akurasi, presisi, dan recall. Penelitian ini juga menyarankan untuk mengecek lebih lanjut pengaruh korelasi antar fitur dan mempertimbangkan penggunaan algoritma lain untuk perbandingan performa klasifikasi.</p>
--	--	--	---

<p>"Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient". Paembonan, S., Abduh, H., dan Kunci, K. (2021).</p>	<p>Metode yang digunakan adalah metode <i>K-Means Clustering</i> dan metode <i>Silhouette Coefficient Clustering</i> untuk evaluasi <i>Clustering</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah pengelompokan obat-obatan yang memiliki kemiripan karakteristik ke dalam kelompok data tertentu menggunakan metode <i>K-Means Clustering</i>. Penelitian ini bertujuan untuk memudahkan deskripsi sifat-sifat atau karakteristik dari masing-masing kelompok obat dengan cara mengelompokkan obat yang mirip ke dalam satu <i>Cluster</i> dan obat yang berbeda ke dalam <i>Cluster</i> yang lain.,</p>	<p>Hasil penelitian ini menunjukkan bahwa metode <i>K-Means Clustering</i> efektif untuk mengelompokkan obat-obatan berdasarkan karakteristiknya. Penelitian ini menemukan bahwa penggunaan dua <i>Cluster</i> adalah yang paling ideal, dengan nilai <i>Silhouette</i> sebesar 0.4854. Hasil ini menunjukkan bahwa pengelompokan obat dengan dua <i>Cluster</i> memberikan pemisahan yang cukup baik antara kelompok obat yang berbeda. Penelitian ini juga menyoroti pentingnya pengelompokan dalam sistem rekomendasi obat dan potensi bagi apoteker</p>
--	--	--	---

			untuk mengganti obat dengan persetujuan dokter dan pasien.
"Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai". Rizuan, R., Haerani, E., Jasril, J., dan Oktavia, L. (2023).	Metode yang digunakan adalah algoritma Mean-Shift,	Masalah yang diangkat dalam penelitian ini adalah untuk menentukan pola karakteristik dari data calon penerima Bantuan Pangan Non Tunai (BPNT) berdasarkan kriteria yang telah ditetapkan oleh Dinas Sosial Kota Pekanbaru.	Hasil penelitian ini menunjukkan bahwa penerapan algoritma Mean-Shift untuk mengelompokkan data penerimaan Bantuan Pangan Non Tunai (BPNT) menghasilkan dua kluster yang berbeda. Kluster 1 terdiri dari penerima dengan status penguasaan tempat tinggal bebas sewa dan

		<p>Penelitian ini menggunakan algoritma <i>Mean-Shift</i> sebagai metode klasterisasi untuk mengidentifikasi pola-pola yang ada dalam data BPNT.</p>	<p>kontrak/sewa, sedangkan Kluster 2 terdiri dari penerima yang memiliki rumah sendiri. Pemilihan bandwidth yang optimal dilakukan menggunakan <i>Silhouette Score</i>, yang menunjukkan bahwa kluster terbaik adalah 2 dengan <i>Bandwidth</i> 285 dan <i>Silhouette Score</i> 0.95. Hasil ini memberikan wawasan tentang pola dan karakteristik penerima BPNT, yang dapat berguna untuk perencanaan program bantuan.</p>
--	--	--	--

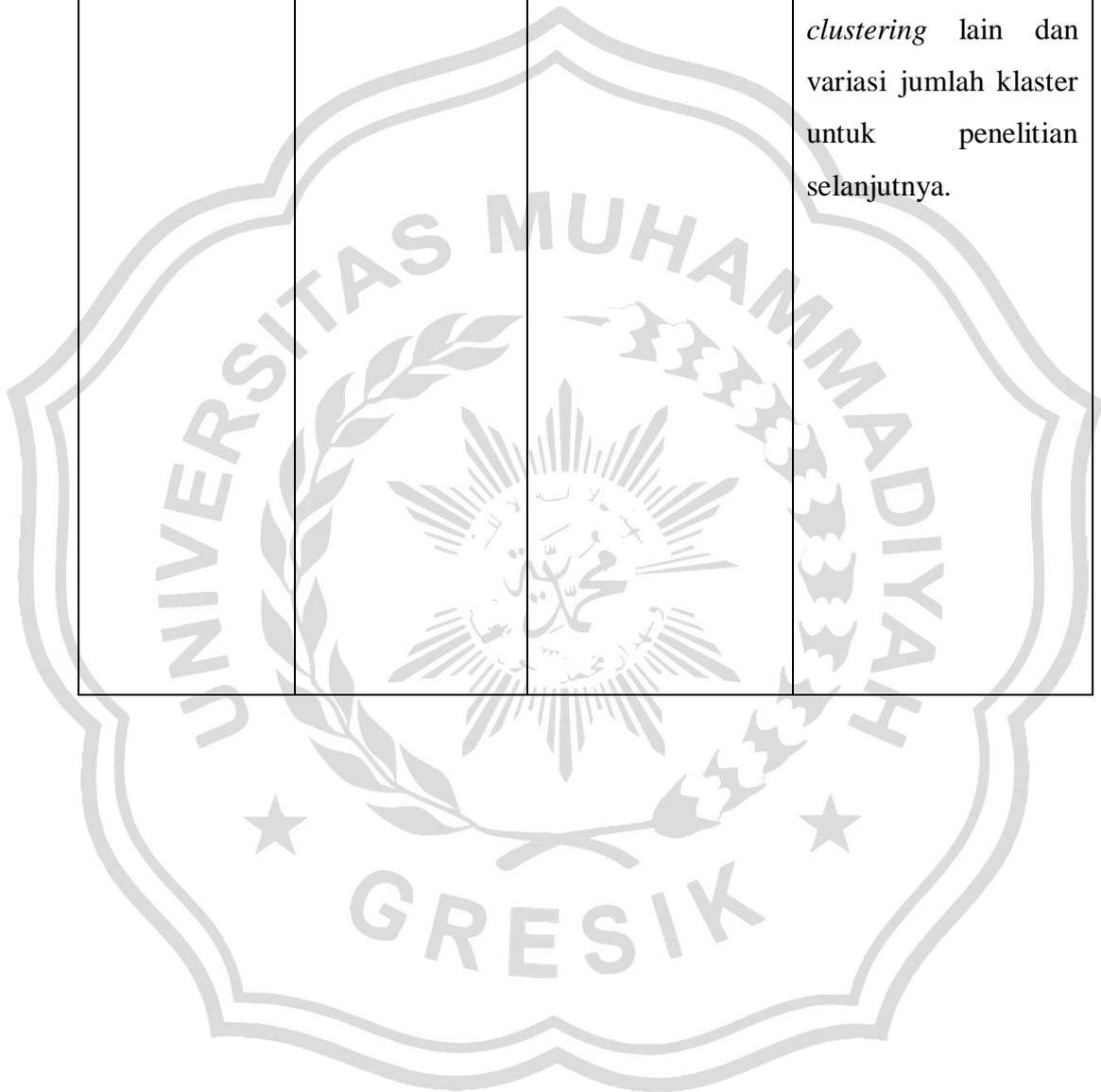
<p><i>"Dampak Gempa Bumi Terhadap Lingkungan Hidup". Sungkawa, D. (2016).</i></p>	<p>Metode yang digunakan untuk menghadapi gempa bumi meliputi persiapan untuk keadaan darurat dan persiapan ketika terjadi gempa bumi. Persiapan untuk keadaan darurat termasuk menentukan tempat berlindung yang aman, menyediakan air minum, dan menyiapkan tas berisi barang-barang penting.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah dampak gempa bumi terhadap lingkungan hidup dan masyarakat, serta tantangan dalam memprediksi dan menghadapi bencana gempa bumi dan tsunami di Indonesia. Gempa bumi sering kali menyebabkan kerusakan besar pada infrastruktur, lingkungan, dan menimbulkan korban jiwa, seperti yang terjadi di Aceh, Sumatera Utara, Yogyakarta, dan Sumatera Barat. Penelitian ini juga menyoroti pentingnya kesiapsiagaan dan mitigasi bencana di</p>	<p>Hasil penelitian ini adalah dalam upaya memitigasi dampak gempa bumi pemerintah pusat dan daerah diharapkan mampu membangun sarana dan prasarana umum, lingkungan kerja dan lingkungan pemukiman yang sesuai dengan daerah gempa. Lembaga-lembaga yang terkait seperti DVMBG yang menangani masalah kegempaan mampu memanfaatkan temuan-temuan di bidang ilmu pengetahuan dan teknologi untuk kepentingan masyarakat. Demikian juga media massa mampu memberikan informasi dan keterangan tentang gempa bumi yang bermanfaat bagi</p>
---	---	---	--

		daerah rawan gempa.	masyarakat. Bagi orang-orang atau penduduk yang tidak mampu beradaptasi di daerah rawan gempa silahkan pindah dan tinggal di daerah yang aman dari gempa, wilayah Indonesia masih luas, jangan tinggal di 28 wilayah rawan gempa dan tsunami menurut DVMBG yang telah di kemukaan oleh penulis pada bagian 3
--	--	---------------------	--

<p>"Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection". Tantyoko, H., Kartika Sari, D., dan Wijaya, A. R. (2023).</p>	<p>Metode yang digunakan dalam penelitian ini adalah algoritma <i>Random Forest</i> dan metode seleksi fitur untuk prediksi gempa bumi di Indonesia.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah prediksi gempa bumi di Indonesia dengan tujuan meningkatkan kemampuan prediksi gempa bumi menggunakan algoritma <i>Random Forest</i> dan metode seleksi fitur. Penelitian ini bertujuan untuk menghasilkan model prediksi yang lebih akurat dan efektif, yang dapat digunakan untuk membantu pihak berwenang dan masyarakat dalam mengambil langkah-langkah mitigasi yang tepat guna mengurangi dampak buruk yang disebabkan oleh gempa bumi.</p>	<p>Hasil penelitian ini menunjukkan bahwa model <i>Random Forest</i>, setelah dilakukan seleksi fitur, mencapai <i>F1 Score</i> sebesar 92,23% dalam memprediksi lokasi gempa bumi di Indonesia. Ini merupakan peningkatan sebesar 5,02% dibandingkan dengan model tanpa seleksi fitur. Selain itu, metode <i>K-Means Clustering</i> digunakan untuk mengelompokkan data gempa menjadi kategori berbahaya dan tidak berbahaya. Hasil ini menekankan pentingnya pemilihan fitur yang tepat untuk meningkatkan performa model prediksi gempa bumi.</p>
--	--	--	--

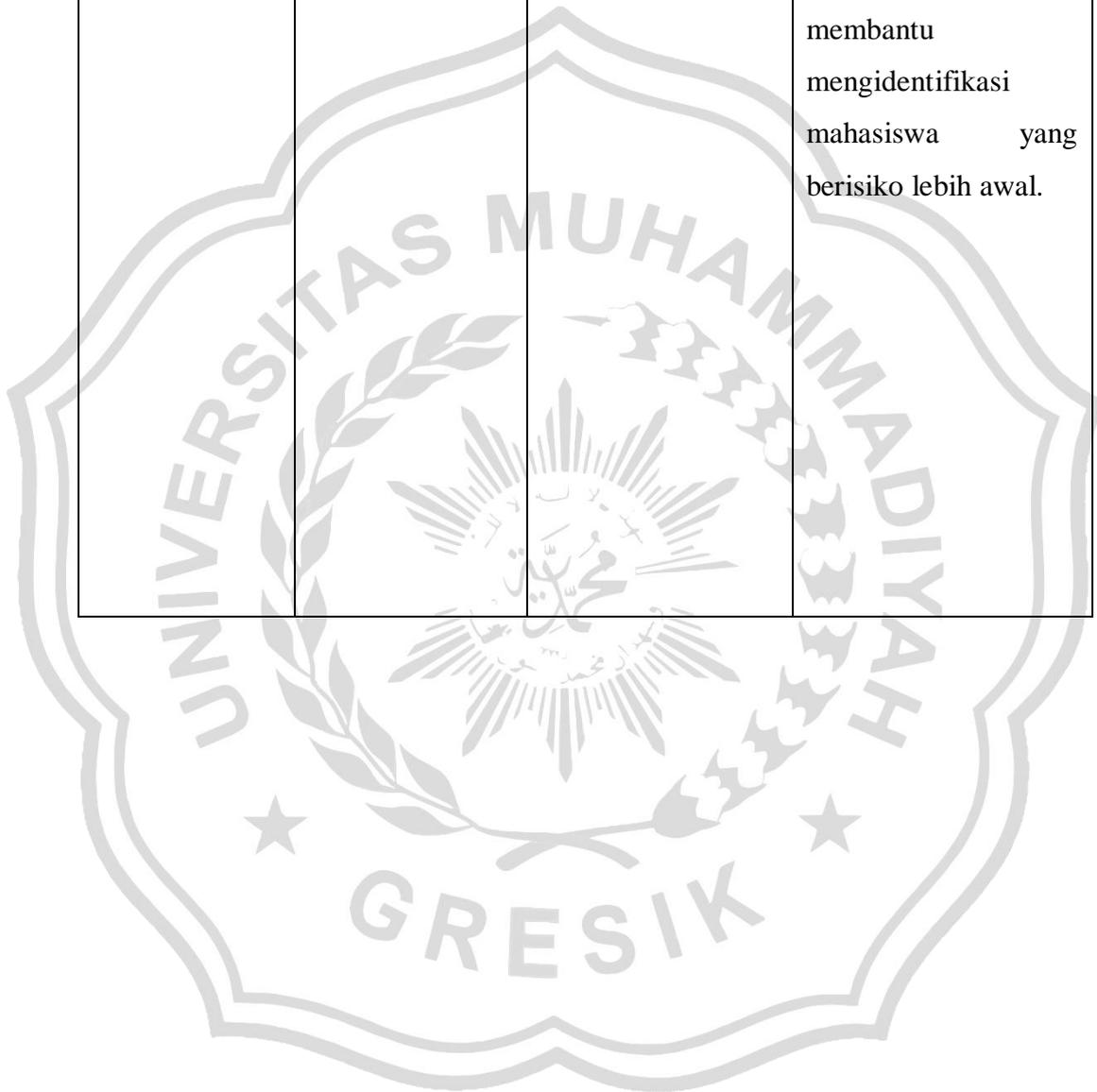
<p><i>"Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means"</i>. Ulinuha, N., Sunan, U., & Surabaya, A. (2020).</p>	<p>Metode yang digunakan adalah <i>Fuzzy C-Means (FCM)</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah pengklasteran produksi komoditas dari sektor perkebunan, terutama kopi, kelapa sawit, biji coklat, dan karet berdasarkan provinsi, dengan tujuan untuk meningkatkan produksi dari komoditas tersebut agar hasil produksi dapat meningkat.</p>	<p>Hasil penelitian ini menunjukkan bahwa provinsi di Indonesia dapat dikelompokkan menjadi tiga klaster berdasarkan produksi komoditas pertanian seperti kelapa sawit, biji karet, biji kopi, dan biji coklat menggunakan metode <i>Fuzzy C-Means</i>. Klaster 1 terdiri dari provinsi Riau yang memiliki produksi tertinggi untuk semua komoditas. Klaster 2 mencakup banyak provinsi dengan produksi terendah, termasuk Aceh dan Sumatera Barat. Klaster 3 terdiri dari provinsi seperti Sumatera Utara dan Kalimantan dengan produksi sedang. Uji <i>Silhouette Index</i> memberikan nilai 0.843, menunjukkan bahwa hasil</p>
---	---	--	---

pengelompokan ini valid. Penelitian ini juga merekomendasikan penggunaan metode *clustering* lain dan variasi jumlah kluster untuk penelitian selanjutnya.



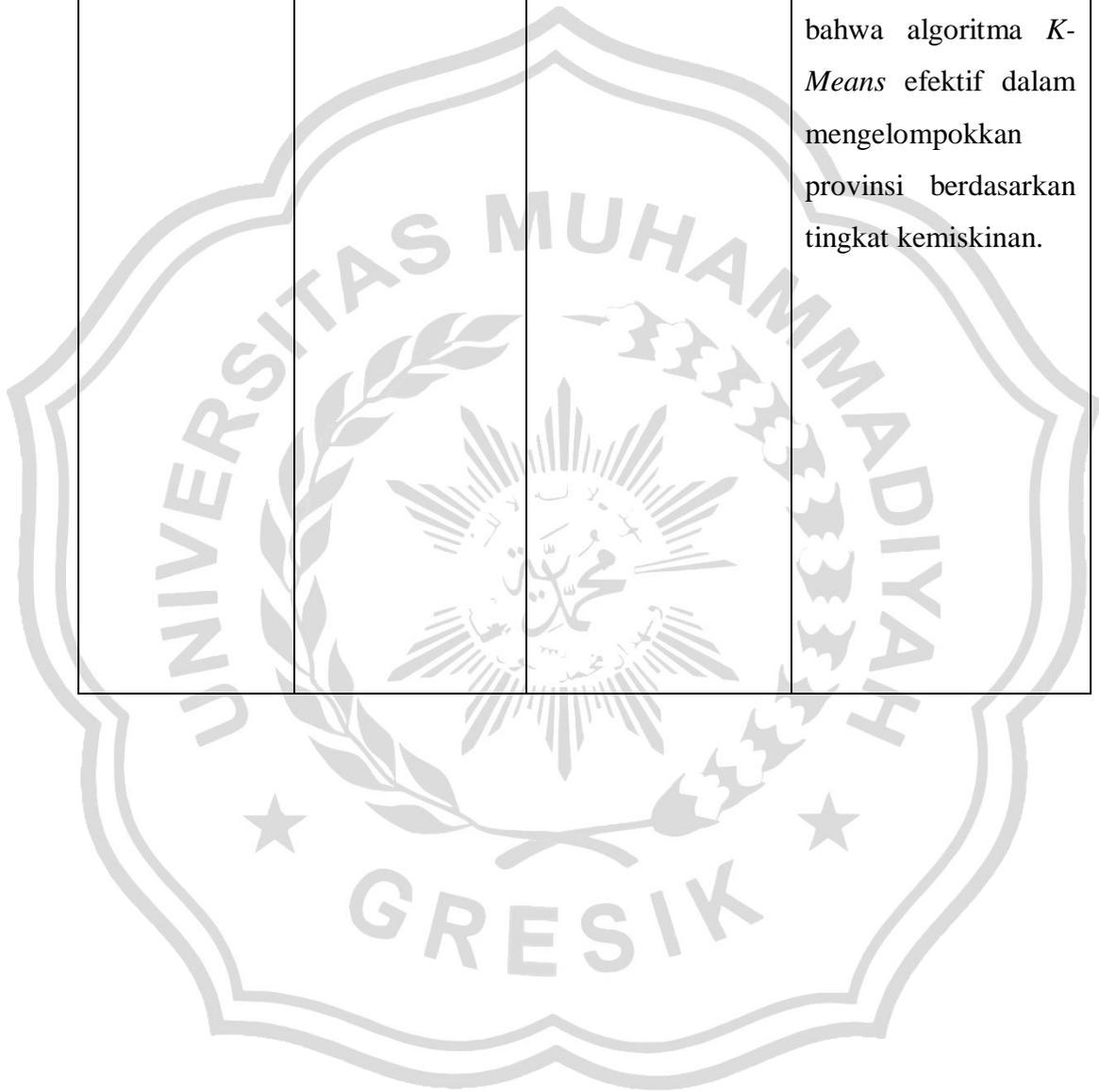
<p><i>"Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, KNN Dan SVM".</i> Widaningsih, S. (2019).</p>	<p>Metode yang digunakan dalam penelitian ini adalah metode <i>Data Mining</i> dengan fungsi klasifikasi. Teknik yang digunakan untuk model data mining klasifikasi terdiri dari empat algoritma yaitu <i>C4.5</i>, <i>Support Vector Machine (SVM)</i>, <i>K-Nearest Neighbor (KNN)</i>, dan <i>Naïve Bayes</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah prediksi nilai dan waktu kelulusan mahasiswa Prodi Teknik Informatika. Penelitian ini bertujuan untuk menggali informasi dari database mahasiswa guna memprediksi waktu kelulusan empat tahun dengan nilai IPK minimal 3,00 menggunakan teknik data mining.</p>	<p>Hasil penelitian ini menunjukkan bahwa algoritma <i>Naïve Bayes</i> adalah yang paling efektif dalam memprediksi kelulusan tepat waktu dan IPK minimal 3,00 untuk mahasiswa Prodi Teknik Informatika di Universitas Suryakencana. <i>Naïve Bayes</i> mencapai akurasi sebesar 76,79%, tingkat kesalahan 23,17%, dan <i>Area Under the Curve (AUC)</i> sebesar 0,850. Algoritma ini mengungguli algoritma lainnya seperti <i>C4.5</i>, <i>K-Nearest Neighbors (KNN)</i>, dan <i>Support Vector Machine (SVM)</i>, yang dikategorikan sebagai cukup dalam kinerja. Penelitian ini menekankan bahwa</p>
--	---	--	---

			teknik <i>Data Mining</i> dapat secara efektif mengidentifikasi pola dalam data kelulusan mahasiswa dan membantu mengidentifikasi mahasiswa yang berisiko lebih awal.
--	--	--	---



<p>"Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means". Yuni Radana Sembiring, Saifullah, dan Riki Winanjaya. (2021).</p>	<p>Metode yang digunakan adalah teknik clustering dengan algoritma <i>K-Means</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah kemiskinan, yang merupakan permasalahan sosial yang dihadapi hampir di setiap negara, termasuk Indonesia. Penelitian ini berfokus pada mengelompokkan jumlah penduduk miskin berdasarkan provinsi menggunakan algoritma <i>K-Means</i> untuk membantu mengidentifikasi faktor-faktor kemiskinan di setiap provinsi dan memberikan informasi atau masukan bagi pemerintah dalam upaya menurunkan jumlah kemiskinan.</p>	<p>Hasil penelitian ini menunjukkan bahwa algoritma <i>K-Means</i> berhasil mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah penduduk miskin menjadi tiga klaster. Klaster 1 terdiri dari 3 provinsi: Jawa Barat, Jawa Tengah, dan Jawa Timur. Klaster 2 terdiri dari 9 provinsi: Aceh, Sumatra Utara, Sumatra Selatan, Lampung, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Selatan, dan Papua. Klaster 3 terdiri dari 22 provinsi, termasuk Sumatra Barat, Riau, Jambi, dan lainnya. Analisis ini dilakukan menggunakan <i>RapidMiner</i> dan hasil klasterisasi dikonfirmasi melalui</p>
---	--	---	--

tiga iterasi yang menghasilkan pembagian kluster yang stabil. Penelitian ini menyimpulkan bahwa algoritma *K-Means* efektif dalam mengelompokkan provinsi berdasarkan tingkat kemiskinan.



<p><i>"Analisis Parameter Gempabumi Dengan Struktur Geologi Di Daerah Asparaga, Gorontalo".</i> Madusila, R. S., Manyoe, I. N., dan Male, C. C. (2021)</p>	<p>Metode yang digunakan dalam penelitian ini meliputi metode seismisitas dengan pembuatan peta seismisitas menggunakan perangkat lunak <i>Quantum Geographical Information System (QGIS)</i> dan metode <i>Peak Ground Acceleration (PGA)</i>.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah analisis gempabumi di daerah Asparaga berdasarkan kedalaman dan magnitudo gempabumi serta hubungannya dengan struktur geologi yang bekerja pada daerah penelitian.</p>	<p>Hasil penelitian ini menunjukkan bahwa daerah Asparaga di Gorontalo, Indonesia, mengalami gempabumi dangkal dengan magnitudo sedang hingga besar, yang erat kaitannya dengan struktur geologi regional. Penelitian ini menemukan bahwa bagian timur Asparaga didominasi oleh gempabumi dengan magnitudo sedang (5.0-5.9 Mw) dan kedalaman dangkal (0-70 km), sementara bagian barat mengalami gempabumi dengan magnitudo lebih tinggi (6.0-6.9 Mw). Studi ini menekankan perlunya survei geologi dan upaya mitigasi untuk meminimalkan</p>
--	---	---	---

dampak gempa bumi
di wilayah tersebut.



<p>"<i>Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means Analisis Earthquake in Java Island Using Clustering K-Means Algorithm</i>". Reviantika, F., Harahap, C. N., dan Azhar, Y. (2020).</p>	<p>Metode yang digunakan dalam penelitian ini adalah algoritma <i>K-Means</i> dan <i>DBSCAN</i></p>	<p>Masalah yang diangkat dalam penelitian ini adalah untuk melihat kawasan yang sering terjadi bencana gempa bumi di Pulau Jawa, sehingga proses mitigasi bencana gempa bumi dapat diterapkan pada kawasan tersebut.</p>	<p>Hasil penelitian ini menunjukkan bahwa metode <i>K-Means</i> menghasilkan dua <i>cluster</i>, sementara metode <i>DBSCAN</i> menghasilkan tiga <i>cluster</i>. Meskipun jumlah <i>cluster</i> berbeda, jumlah data dalam masing-masing <i>cluster</i> relatif sama. Indeks <i>Silhouette</i> menunjukkan bahwa <i>K-Means</i> lebih unggul dibandingkan <i>DBSCAN</i> dalam hal kualitas <i>clustering</i>, dengan nilai indeks <i>Silhouette</i> 0.54 untuk <i>K-Means</i> dibandingkan 0.17 untuk <i>DBSCAN</i>. Penelitian ini menyoroti pentingnya teknik data mining dalam memahami dan mengelola risiko gempa bumi di daerah yang padat penduduk.</p>
---	---	--	--

<p>"Peningkatan Performa Cluster Fuzzy C-Means Pada Pengklasteran Sentimen Menggunakan Particle An Improved Fuzzy C-Means For Sentiment Clustering Based On Particle Swarm Optimization". Siringoringo, R., & Jamaludin. (2019).</p>	<p>Metode yang digunakan adalah <i>Fuzzy C-Means (FCM)</i> yang ditingkatkan dengan penerapan <i>Particle Swarm Optimization (PSO)</i> untuk pengklasteran sentimen.</p>	<p>Masalah yang diangkat dalam penelitian ini adalah sensitifitas <i>Fuzzy C-Means (FCM)</i> terhadap pusat klaster, yang dapat menghasilkan iterasi proses yang sangat rumit dan mengakibatkan proses pengklasteran terjebak pada kondisi optimum lokal. Sensitifitas ini membuat hasil klasterisasi sulit dikontrol dan sangat tergantung pada inisialisasi <i>centroid</i>.</p>	<p>Hasil penelitian ini menunjukkan bahwa metode <i>Fuzzy C-Means</i> yang dioptimalkan dengan <i>Particle Swarm Optimization (FCM-PSO)</i> memberikan hasil yang lebih baik dibandingkan dengan <i>FCM</i> konvensional. Peningkatan ini diukur berdasarkan kriteria seperti <i>Rand Index</i>, <i>F-Measure</i>, dan <i>Objective Function Value (OFV)</i>. <i>FCM-PSO</i> memiliki waktu konvergensi yang lebih cepat dan penanganan noise yang lebih baik, terutama dalam data yang berdimensi tinggi dan tidak terstruktur. Penelitian ini menggunakan data ulasan produk dari situs belanja online di</p>
--	--	--	---

Indonesia untuk
analisis sentimen.

