G-Tech : Jurnal Teknologi Terapan
Volume 7, No. 3, Juli 2023, hal. 1218 - 1226
E-ISSN: 2623-064x | P-ISSN: 2580-8737

Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma K-Means

Ilsa Hidayat[1]✉, Eva Darnila[2], Yesy Afrillia[3]

[1, 2, 3] Informatika, Fakultas Teknik, Universitas Malikussaleh, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 09-07-2023
Direvisi : 15-07-2023
Diterima : 16-07-2023

ABSTRAK

Indonesia sering mengalami bencana alam, khususnya banjir di Panyabungan, Kabupaten Mandailing Natal, menghadapi tantangan besar dalam menjaga keselamatan dan kesejahteraan penduduknya. Tingginya kerentanan terhadap bencana alam dan kurangnya pemetaan yang akurat terhadap daerah rawan banjir menjadi kendala dalam penanggulangan bencana. Namun, melalui penggunaan teknologi data mining dan algoritma K-Means, penelitian ini menawarkan solusi potensial untuk secara efektif mengidentifikasi wilayahyang rentan terhadap banjir.Pengembangan model pengelompokan dan perangkat lunak ini akan membantu pihak berwenang untuk mengurangi kerugian akibat bencana banjir. Dengan pemetaan daerah rawan dan pengklasteran menjadi tiga tingkat kerentanan, respons terhadap bencana dapat dilakukan dengan lebih cepat dan efisien. Penelitian ini bertujuan untuk mendukung upaya mitigasi bencana dan peningkatan pelayanan bagi masyarakat di Panyabungan. Dengan demikian, penelitian ini memberikan kontribusi penting dalam meningkatkan pemahaman tentang bencana banjir di wilayah rentan di Indonesia, serta bisa menjadi acuan untuk pengambilan keputusan yang tepat guna menjaga keselamatan penduduk di masa depan.

Kata Kunci:

Data Mining, Clustering, Banjir.

***Keywords* :**

Data Mining, Clustering, Flooding.

ABSTRACT

Indonesia often experiences natural disasters, particularly floods in Panyabungan, Mandailing Natal Regency, facing significant challenges in safeguarding the safety and well-being of its population. The high vulnerability to natural disasters and the lack of accurate mapping of flood-prone areas pose obstacles in disaster management. However, through the use of data mining technology and the K-Means algorithm, this research offers a potential solution to effectively identify flood-prone areas. The development of this clustering model and software will assist authorities in reducing losses caused by flood disasters. By mapping vulnerable areas and clustering them into three levels of vulnerability, disaster response can be carried out more quickly and efficiently. This research aims to support disaster mitigation efforts and improve services for the community in Panyabungan. Thus, this study provides an important contribution to enhancing understanding of flood disasters in vulnerable regions in Indonesia and can serve as a reference for making informed decisions to ensure the safety of the population in the future.

***Corresponding Author* :**
Ilsa hidayat
Informatika, Fakultas Teknik, Universitas Malikussaleh
Blang Pulo, Muara Satu, Kota Lhokseumawe
Email: ilsahidayat248@gmail.com

This is an open access article under the CC BY license

PENDAHULUAN

Indonesia sering dilanda bencana alam yang mengancam keberlangsungan hidup manusia, seperti banjir. Topografi Indonesia yang memiliki banyak daerah pegunungan menjadi faktor terjadinya bencana alam di berbagai tempat. Dampak bencana alam sangat beragam, termasuk kehilangan harta benda, kerusakan tempat tinggal, dan pemisahan keluarga. Namun, dampak tersebut dapat diminimalisir dengan peran teknologi dan manajemen bencana yang baik.

Kabupaten Mandailing Natal di Provinsi Sumatera Utara berbatasan dengan Sumatera Barat dan memiliki topografi gugusan pegunungan dan perbukitan, termasuk bukit Barisan dan gunung berapi aktif gunung Sorik Marapi. Kecamatan Panyabungan di Kabupaten Mandailing Natal memiliki topografi curam dan curah hujan tinggi, membuatnya rentan terhadap bencana alam, khususnya banjir. Selama periode 2020-2021, Kecamatan Panyabungan sering dilanda bencana banjir di banyak wilayah seperti Aek Banir, Sipaga-Paga, Parbangunan, dan banyak lainnya. Meskipun kasus bencana alam di Kecamatan Panyabungan terus meningkat, sulit untuk menentukan daerah yang terkena bencana karena data yang perlu dianalisis cukup banyak. Di kantor BPBD Mandailing Natal, sistem pelayanan pengaduan bencana masih menggunakan metode konvensional dengan pelaporan melalui telepon oleh masyarakat. Namun, kantor BPBD mengalami kesulitan dalam mengolah informasi untuk disampaikan kepada masyarakat Panyabungan.

Data mining adalah bagian dari Big Data, yang telah menjadi salah satu bahasa yang paling populer di industri. Big Data merupakan konsep abstrak yang umumnya digambarkan sebagai integrasi data yang sangat besar dan kompleks, sulit diolah dengan alat manajemen database tradisional. Big Data sering ditandai dengan lima karakteristik utama: kuantitas, kecepatan, keragaman, nilai, dan akurasi (Yang et al., 2020).

Jika kita membahas *big data* istilah data mining juga tidak akan terlepas dari *big data*. Menerapkan teknologi data mining dapat dengan mudah dalam menganalisa data seperti data daerah rawan banjir. penerapan teknik data mining yang dapat kita gunakan untuk mempermudah dalam menganalisa data dengan *clustering* yaitu Algoritma *K-Means* (Sukma et al., 2019). Algoritma *K-Means* termasuk metode sederhana dan cepat dalam melakukan proses *clustering*. Dengan Algoritma *K-Means* dapat mempermudah untuk membuat penentuan daerah mana saja yang terkena bencana banjir dan yang tidak terkena bencana banjir. Pekerjaan yang dilakukan akan lebih mudah dan optimal dengan menggunakan Algoritma *K-Means* (Nisar et al., 2022).

Penulis mengatasi permasalahan tersebut dengan membuat model Clustering menggunakan algoritma K-Means. Model ini dirancang dan diimplementasikan melalui perangkat lunak berbasis Web untuk membuat pemetaan zonasi rawan bencana. Pemetaan ini berfungsi sebagai acuan dalam mengantisipasi bencana banjir di masa depan dan mendukung kecepatan layanan informasi musibah di Kecamatan Panyabungan. Dalam pengklasteran menggunakan algoritma K-Means, penulis menggunakan tiga kriteria: Sangat Rawan, rawan, dan tidak rawan (aman dari bencana.

METODE PENELITIAN

Pendekatan penelitian adalah serangkaian langkah atau metode yang dilakukan oleh peneliti untuk memperoleh informasi atau data yang diperlukan. Berikut ini adalah prosedur yang penulis terapkan dalam penelitiannya antara lain: 1) pengumpulan data; 2) studi keputusan: 3) wawancara 4) implementasi; 5) pengujian sistem

1) Data *Mining*
   Data *mining* menerapkan dua metode langkah, yaitu Unsupervised learning dan Supervised learning. Unsupervised learning tidak menggunakan bimbingan atau instruktur, tetapi menggunakan nama data sebagai guru. Sedangkan Supervised learning melibatkan bimbingan dan instruktur.(Enda Esyudha Pratama et al., 2021)

2) Clustering

Clustering merupakan Proses yang di gunakan untuk melibatkan perancangan anggota sehingga seluru anggota setiap kelompok mempunyai kesamaan terhadap matriks tertentu.(Ananda et al., 2022)

3) Metode Clustering K-Means

melibatkan cara iteratif untuk prosesnya. Dalam konteks K-Means, huruf K mewakili banyak cluster yang akan dibentuk. Seterusnya, nilai K ditentukan secara acak/random sebagai inisialisasi. Di sisi lain, "means" merupakan nilai tengah yang berfungsi sebagai pusat dari setiap klaster, juga dikenal sebagai centroid (Harahap et al., 2022). Pada prosedurnya Metode k-means dapat di jabrkan di bawah ini :

a. Tentukan brapa banyak Cluster (k) terhadapa data set.

b. Menginisialisasi titik pusat (Centroid) secara acak.

c. Menggunakan metode perhitungan jarak yang paling dekat dengan centroid, dapat menggunakan rumus (1) di bawah ini:

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Keterangan :
- d = Euclidean Distance.
- i = banyak objek.
- x,y = Titik koordinat objek.
- s,t = Titik koordinat centroid.

d. Lakukan cara pengelompokan objek berlandaskan jarak yang paling dekat dengan centroid.

e. Hitung rata-rata untuk setiap kelompok menggunakan rumus (2) di bawah:

$$V_{ij} = \frac{1}{N_i}\sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Keterangan :
- V_ij = Centroid rata-rata cluster ke-i untuk variable ke-j.
- N_i = Jumlah anggota cluster ke-i.
- i,k = Indeks dari cluster.
- j = Indeks dari variable.
- X_kj = Nilai data ke-k variable ke-j untuk cluster tersebut.

f. Lakukan perulanagan dari poin 3 serta poin 4 dan iterasi hingga mencapai centroid dengan nilai optimal.(Risawandi & Afrillia, 2022)

4) Sistem Informasi Geografis

Sistem Informasi Geografis atau bisa di singkat dengan (SIG) adalah sebuag sistem yang menyampaikan informasi berbasis computer dengan menggunakan data dengan komponen informasi spasial.(Jannah et al., 2022)

5) PHP

Mulanya PHP dipakai sebagai bahasa pemrograman buat server-side HTML-embedded yang di promosikan sebagai Individu Home Pages. Rasmus Lerdorf merupakan pembuat sebuah Bahasa pemograman yaitu PHP waktu masa 1995. Pada masa itu, PHP di juluki FI (Form Interpreted).(Risawandi, 2019)

HASIL DAN PEMBAHASAN

Pengujian perhitungan program sistem yang telah dibangun dilakukan melalui tes uji perhitungan secara manual. Tujuan dari tes ini adalah untuk mendapatkan hasil perhitungan sistem yang sama dengan perhitungan di luar sistem menggunakan algoritma k-means. Pengujian ini menggunakan data tentang daerah rawan banjir di setiap Desa di Kecamatan Panyabungan

yang di peroleh dari BPBD. Dalam proses clustering, data tersebut akan dihitung berdasarkan jumlah data mengenai daerah rawan banjir dari tahun 2019 hingga 2022 untuk 39 Desa di Wilayah Kecamatan Panyabungan. Data ini merupakan data awal sebelum dilakukan normalisasi.

Tabel 1. Tabel Data awal Tahun 2019

No	Nama Desa	Jumlah Kejadian Banjir	Jumlah Korban Jiwa Banjir	Jumlah Bangunan rumah di bantaran sungai
1	Aek Banir	3	3	41
2	Sipaga - Paga	1	1	11
….	…….	…….	……	……
38	Panggorengan	0	0	0
39	Manyabar Jae	2	5	45

Tabel 2. Tabel Data awal Tahun 2020

No	Nama Desa	Jumlah Kejadian Banjir	Jumlah Korban Jiwa Banjir	Jumlah Bangunan rumah di bantaran sungai
1	Aek Banir	2	4	43
2	Sipaga - Paga	2	2	23
….	…….	…….	……	……
38	Panggorengan	0	0	0
39	Manyabar Jae	3	3	46

Tabel 3. Tabel Data awal Tahun 2021

No	Nama Desa	Jumlah Kejadian Banjir	Jumlah Korban Jiwa Banjir	Jumlah Bangunan rumah di bantaran sungai
1	Aek Banir	3	3	45
2	Sipaga - Paga	1	2	15
….	…….	…….	……	……
38	Panggorengan	0	0	0
39	Manyabar Jae	3	3	45

Tabel 4. Tabel Data awal Tahun 2022

No	Nama Desa	Jumlah Kejadian Banjir	Jumlah Korban Jiwa Banjir	Jumlah Bangunan rumah di bantaran sungai
1	Aek Banir	3	4	47
2	Sipaga - Paga	2	2	23
….	…….	…….	……	……
38	Panggorengan	0	0	0
39	Manyabar Jae	4	3	44

Adapun langkah-langkah yang dilakukan dalam menyelesaikan perhitungan Algoritma K-Means adalah sebagai berikut:

Normalisasai Data

Normalisasi data menguunakan rumus di bawah ini.(Ahmad Harmain et al., 2022)

$$Data\ Kriteria = \frac{Data\ awal - Min(Data\ Kriteria)}{Max(Data\ Kriteria) - Min(Data\ Kriteria)} \quad (3)$$

Tabel 5. Tabel Hasil Normalisasi

No	Nama Desa	Jumlah Kejadian Banjir	Jumlah Korban Jiwa Banjir	Jumlah Bangunan rumah di bantaran sungai
1	Aek Banir	0,75	0,5	0,891304
2	Sipaga - Paga	0,25	0,166667	0,23913
3	Parbangunan	0,25	0,333333	0,23913
4	Pidoli Lombang	0,25	0,333333	0,23913
5	Pidoli Dolok	0,5	0,666667	0,76087
....	……	…….	……	…….
....	…….	…….	……	……
37	Salambue	0	0	0
38	Panggorengan	0	0	0
39	Manyabar Jae	0,5	0,833333	0,978261

Menentukan Jumlah Cluster

Ada 3 cluster yang digunakan dalam analisis tersebut, yaitu Sangat Rawan (C1), Rawan (C2), dan Tidak Rawan (C3), berdasarkan data banjir yang diperoleh dari BPBD selama empat tahun terakhir (2019-2022), dengan total jumlah desa sebanyak 39.

Tabel 6. Jumlah Klaster Potensi

Klaster	Label Potensi
C1	Sangat Rawan
C2	rawan
C3	Tidak Rawan

Menetapkan Centroid awal

Untuk Menetapkan titik pusat awal klaster centroid, dilakukan dengan cara manual yang diperoleh dari data penelitian. Penentuan titik pusat awal ataupun centroid awal dilakukan dengan cara membentuk 3 kelompok, setelah itu pilih lah 3 data yang dijadikan centroid awal secara acak ataupun random. Penentuan centroid awal ini di ambil dari data yang telah di normalisasi.

Tabel 7. Centroid Awal yang Telah di Tentukan

Label Potensi	x	y	z
C1	0.75	0.5	0.891304
C2	0.25	0.166667	0.23913
C3	0	0	0

Proses Hitung Iterasi Awal

Langkah yang di lakukan yaitu :

1. Lakukan Perhitungan manual Jarak Euclidean
   $C1.1=\sqrt{((0.75-0.75)^2+(0.5-0.5)^2+(0.891304-0.891304)^2)}=0$
   $C2.1=\sqrt{((0.75-0.25)^2+(0.5-0.166667)^2+(0.891304-0.23913)^2)}=0.886816$
   $C3.1=\sqrt{((0.75-0)^2+(0.5-0)^2+(0.891304-0)^2)}=1.267645$

2. Menentukan nilai terkecil ataupun nilain minimum dari keseluruhan jarak klaster yang sudah dihitung. dan kemudian dijabarkan dalam bentuk tabel dibawah ini.

Tabel 8.Hasil Jarak Euclidean awal

No	Nama Desa	C1	C2	C3	Jarak Terkecil	Label Potensi	Predikat
1	Aek Banir	0	0,886816	1,267645	0	C1	Sangat Rawan
2	Sipaga - Paga	0,886816	0	0,384007	0	C2	Rawan
3	Parbangunan	0,838516	0,166666	0,48041	0,166666	C2	Rawan
4	Pidoli Lombang	0,838516	0,166666	0,48041	0,166666	C2	Rawan
5	Pidoli Dolok	0,327553	0,764665	1,128436	0,327553	C1	Sangat Rawan
…..	…..	…	….	…	….	….	….
….	…..	….	…..	….	….	….	…..
37	Salambue	1,267645	0,384007	0	0	C3	Tidak Rawan
38	Aek Mata	1,267645	0,384007	0	0	C3	Tidak Rawan
39	Huta Lombang Lubis	1,267645	0,384007	0	0	C3	Tidak Rawan

3. Langkah selanjutnya adalah menetapkan centroid baru untuk melakukan perhitungan nilai dengan menjumlahkan dari data yang sudah dinormalisasi pada variabel yang memiliki hasil klaster yang sama. Prosedur ini diaplikasikan atau diterapkan pada C1, C2, dan C3, kemudian dibagi dengan jumlah kelompok dalam masing-masing klaster C1, C2, dan C3.

$$Centroid\ Baru = \frac{nilai\ rata{-}rata\ cluster\ 1}{jumlah\ data\ cluster\ 1} \quad (4)$$

Cluster 1
C1 =(0.75+0.5+0.75+0.5+……..1+0.5)÷13=0.596154
C1 =(0.5+0.666667+0.333333+……+0.833333)÷13=0.666667
C1 =(0.891304+……+0.913043+0.978261)÷13= 0.827759
Cluster 2
C2 =(0.25+0.25+0.25+0.25+0.25+0.25+0.25+0.25)÷8=0.25
C2 =(0.166667+0.333333+…….+0.166667)÷8=0.25
C2 =(0.23913+0.23913+…..+0.26087)÷8 =0.222826
Cluster 3
C3 =(0+0+0+0+0+0+….+0+0+0+0+0+0+0)÷180=0
C3 =(0+0+0+0+0+…….+0+0+0+0+0+0+0+0+0)÷18=0
C3 =(0+0+0+0+0+0+0+……+0+0+0+0+0+0+0)÷18=0
Berikut ini adalah tabel yang menampilkan hasil perhitungan yang telah di hitung dari pencarian centroid terbaru:

Tabel 9. Centroid Baru Iterasi 1

Label Potensi	x	y	z
C1	0.596154	0.666667	0.827759
C2	0.25	0.25	0.222826
C3	0	0	0

Proses Hitung Iterasi Ke-2

Pada tahapan Iterasi kedua ini melibatkan centroid yang digunakan dalam perhitungan sebelumnya. Nilai centroid baru setelah proses Euclidean awal juga ditampilkan dalam tabel di bawah ini. Centroid baru yang telah di hitung ini digunakan untuk melakukan perhitungan lanjutan dengan menggunakan metode Euclidean distance sampai menemukan kesamaan antara centroid sebelumnya. Hasil perhitungan tersebut ditunjukkan pada Tabel 10 dengan menggunakan data yang telah dinormalisasi.

Tabel 10.Hasil Jarak Euclidean ke-2

No	Nama Desa	C1	C2	C3	Jarak Terkecil	Label Potensi	Predikat
1	Aek Banir	0.235551	0.871414	1.267645	0.235551	C1	Sangat Rawan
2	Sipaga - Paga	0.846349	0.084913	0.384007	0.084913	C2	Rawan
3	Parbangunan	0.75988	0.084913	0.48041	0.084913	C2	Rawan
4	Pidoli Lombang	0.75988	0.084913	0.48041	0.084913	C2	Rawan
5	Pidoli Dolok	0.117131	0.724985	1.128436	0.117131	C1	Sangat Rawan
…..	…..	…	….	…	….	….	….
….	…..	….	…..	….	….	….	…..
37	Salambue	1.218618	0.417913	0	0	C3	Tidak Rawan
38	Aek Mata	1.218618	0.417913	0	0	C3	Tidak Rawan
39	Huta Lombang Lubis	0.244283	0.98664	1.378927	0.244283	C1	Sangat Rawan

Mencari nilai centroid baru dilakukan dengan cara menambahakan keseluruhan kelompok cluster dari data, kemudian dibagi dengan jumlah kelompok dalam masing-masing klaster, Berikut ini adalah hasil yang terdapat dalam tabel 12 centroid yang baru:

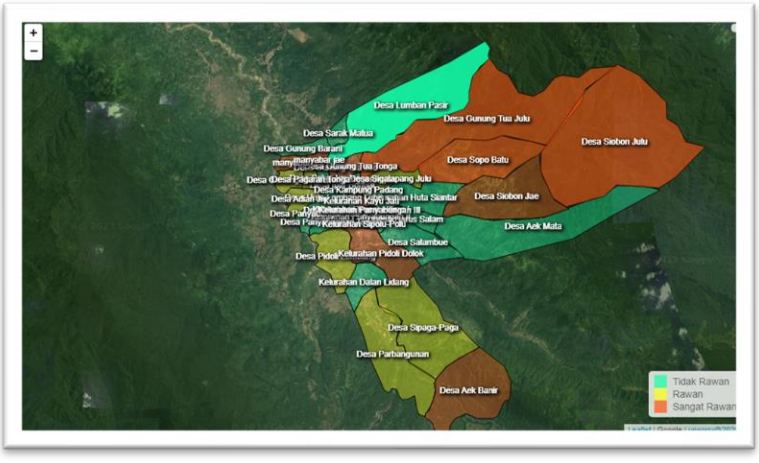
Tabel 11. Centroid Baru Iterasi 2

Label Potensi	x	y	z
C1	0.596154	0.666667	0.827759
C2	0.25	0.25	0.222826
C3	0	0	0

Sesudah menemukan centroid baru, langkah yang harus di lakukan selanjutnya adalah membandingkan dengan centroid sebelumnya. apabila kedua centroid tersebut berbeda, proses akan dilanjutkan ke iterasi selanjutnya sampai nilai centroidnya ada yang sama. Namun, jika ada kedua centroid tersebut sudah memiliki nilai yang sama, maka proses akan dihentikan. Dari Proses iterasi ke-1 dan ke-2 sudah sama maka proses di hentikan.

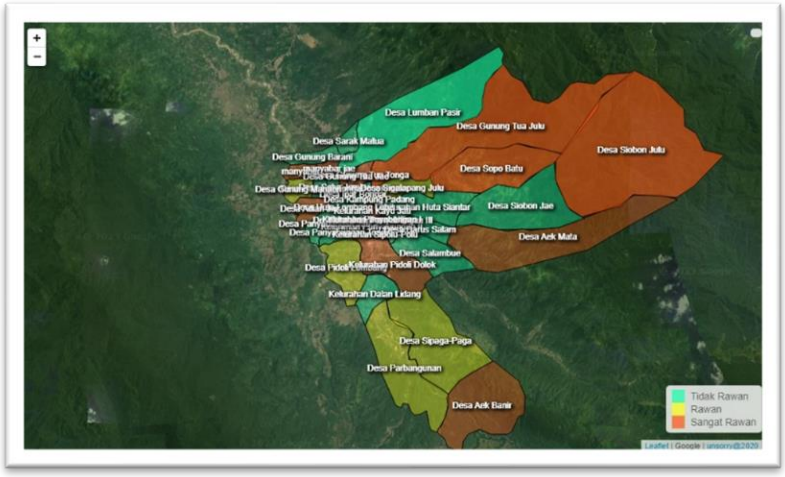
Implementasi Syistem Hasil Pemetaan

Hasil clustering Zonasi Bencana Banjir Kecamatan Panyabungan Tahun 2019 sampai 2022. Berikut gambar Peta hasil klastering zonasi Bencana banjir disetiap desa:



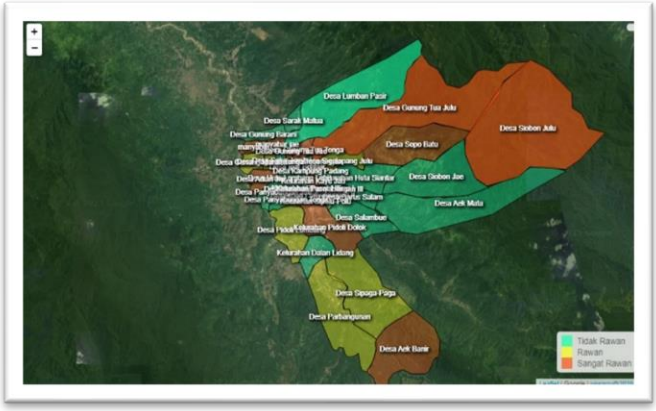
Gambar 1. Peta Rawan Banjir Tahun 2019

Jumlah desa dari hasil potensitensi tahun 2019 yaitu sangat rawan 13,rawan 8 dan tidak rawan 18.



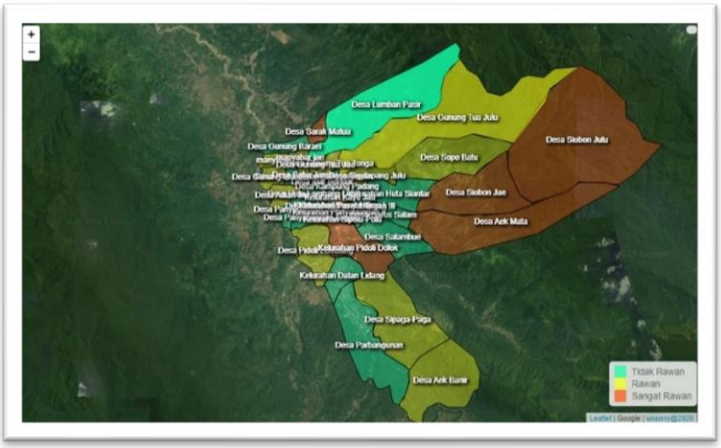
Gambar 2. Peta Rawan Banjir Tahun 2020

Jumlah desa dari hasil potensitensi tahun 2020 yaitu sangat rawan 14,rawan 7 dan tidak rawan 18.



Gambar 3. Peta Rawan Banjir Tahun 2021

Jumlah desa dari hasil potensitensi tahun 2021 yaitu sangat rawan 13,rawan 7 dan tidak rawan 19.



Gambar 4. Peta Rawan Banjir Tahun 2022

Jumlah desa dari hasil potensitensi tahun 2022 yaitu sangat rawan 7, rawan 14 dan tidak rawan 18.

KESIMPULAN DAN SARAN

Kesimpulan

Algoritma K-Means Klastering telah berhasil diterapkan untuk menentukan daerah rawan banjir dalam 3 klaster: sangat rawan, rawan, dan tidak rawan. Outputnya berupa sistem berbasis web. Hasil klastering banjir berdasarkan tahun 2019, 2020, 2021, dan 2022 menunjukkan desa-desa yang masuk ke dalam kategori sangat rawan, rawan, dan tidak rawan banjir.Beberapa desa yang masuk dalam kategori sangat rawan banjir adalah Aek Banir, Pidoli Dolok, Siobon Julu, Panyabungan III, Pasar Hilir, dan lain-lain.Beberapa desa yang masuk dalam kategori rawan banjir termasuk Sipaga-Paga, Parbangunan, Pidoli Lombang, dan lain-lain.Beberapa desa yang tidak rawan banjir termasuk Darussalam, Kota Siantar, Panyabungan II, dan lain-lain..

Saran

Sistem klastering ini masih sangat harus di kembangkan lagi dengan menanbak kriteria banjir. Karena di Kecamatan Panyabungan bukan Cuma banjir saja yang terjadi maka perlu di buat penambahan bencana seperti Longsor dan gempa pumi. Disarankan juga menggunkan menggunakan Metode klastering yang lain untuk menambah variasi dalam sebuah peneltian.

REFERENSI

Ahmad Harmain, Paiman, P., Kurniawan, H., Kusrini, K., & Dina Maulina. (2022). Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, *2*(2), 83–89. https://doi.org/10.46764/teknimedia.v2i2.49

Ananda, W., Santi, I. H., Kirom, S., Informasi, F. T., Balitar, U. I., Blitar, K., & Mining, D. (2022). Penerapan Algoritma K-Means Clustering dalam Pengelompokan Arsip SKCK. *6*(2), 861–867. https://doi.org/https://doi.org/10.36040/jati.v6i2.5762

Enda Esyudha Pratama, Sastypratiwi, H., & Yulianti. (2021). Analisis Kecenderungan Informasi Terkait Covid-10 Berdasarkan Big Data Sosial Media dengan Menggunakan Metode Data Mining. *Jurnal Informatika Polinema*, *7*(2), 1–6. https://doi.org/10.33795/jip.v7i2.453

Harahap, L. M., Fuadi, W., Rosnita, L., Darnila, E., & Meiyanti, R. (2022). Klastering Sayuran Unggulan menggunakan Algoritma K-Means. *Jurnal Teknik Informatika Dan Sistem Informas*, *8*(3), 567–579. https://doi.org/10.28932/jutisi.v8i3.5277

Jannah, M., Muthmainnah, M., Safwandi, S., Saptari, M. A., Muhammad, M., Wahyudi, R., & Farhan, M. (2022). *Implementation of Geographic Information System for Tourist Locations and Lodging Services in Lhokseumawe City Based on Android*. *International Journal of Engineering, Science and Information Technology*, *2*(4), 39–47. https://doi.org/10.52088/ijesty.v2i4.320

Nisar, Wasilah, & Kusumajaya, H. (2022). Pemanfaatan K Means Clustering dalam Pengelompokan Judul Skripsi. *Jurnal JUPITER*, *14*(1), 19–26.

Risawandi. (2019). Mudah Menguasai PHP & MySQL. In *Unimal Press*. UNIMAL PRESS.

Risawandi, & Afrillia, Y. (2022). *Geographic Information System Mapping Of Criminality Villed Areas In Lhokseumawe Using K-Means Method*. *Journal of Informatics and Telecommunication Engineering*, *5*(2), 442–451. https://doi.org/10.31289/jite.v5i2.6265

Sukma, A. R., Halfis, R., & Ady Hermawan. (2019). Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5. *Jurnal Teknik Komputer AMIK BSI*, *4*(2), 21–28. https://doi.org/10.31294/jtk.v4i2

Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., Zheng, S., Xu, A., & Lyu, J. (2020). *Brief introduction of medical database and data mining technology in big data era*. *Journal of Evidence-Based Medicine*, *13*(1), 57–69. https://doi.org/10.1111/jebm.12373