



Comparison Of K-Means And K-Medoid Algorithms For Clustering Of Flood-Prone Areas In Rokan Hilir District

Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir

Ena Tasia^{1*}, M.Afdal²

^{1,2} Program Studi Sistem Informasi, Fakultas Sains dan Teknologi UIN Sultan Syarif Kasim
E-Mail: ¹12050321658@students.uin-suska.ac.id, ²m.afdal@uin-suska.ac.id

Makalah: Diterima 05 Januari 2023; Diperbaiki 23 Februari 2023; Disetujui 03 Maret 2023
Corresponding Author: Ena Tasia

Abstrak

Kabupaten Rokan Hilir memiliki 18 kecamatan dan 156 desa, wilayah ini merupakan daerah yang dilalui aliran sungai atau yang disebut Daerah Aliran Sungai (DAS) yaitu sungai rokan. Sesuai dengan kondisi topografis dan morfologi sungai rokan akan mengalami banjir tahunan yang disebabkan oleh beberapa factor yaitu aktifitas hujan, DAS, wilayah, struktur geologis, aliran air (drainase), serta faktor lainnya yang disebabkan oleh aktifitas manusia yang sering membuka lahan yang menyebabkan daerah ini rawan terhadap banjir. Sejak tahun 2005 banjir menjadi permasalahan di Kabupaten Rokan Hilir hingga saat ini, banjir terparah pernah terjadi pada tahun 2011-2014 yang menyebabkan banyak daerah dan warga secara langsung terdampak akibat hal tersebut, sehingga masyarakat sangat membutuhkan bantuan dari beberapa pihak terutama pihak instansi pemerintahan seperti dinas sosial. Agar dinas social dapat memberikan bantuan yang tepat waktu, lokasi, sasaran, kualitas, kuantitas, dan sesuai kebutuhan masyarakat, dibutuhkan klastering pada daerah rawan banjir. Pada penelitian ini dikelompokkan wilayah yang rawan banjir di kabupaten Rokan Hilir. Untuk melakukan klastering pada penelitian ini akan digunakan algoritma K-Means dan K-Medoids. Setelah dilakukan percobaan menggunakan tools RapidMiner pada klaster K=2 sampai K=6 didapat metode K-Means lebih optimal dibandingkan dengan menggunakan metode K-Medoids pada data kejadian banjir di Rokan Hilir pada tahun 2019-2022 dengan jumlah k paling optimal adalah k = 3 dengan nilai validitas *Davies-Bouldin Index* (DBI) sebesar 0.218%, sedangkan algoritma K-Medoids klaster optimal terdapat pada K=4 dengan nilai validitas sebesar 0.525%.

Kata Kunci: *banjir, davies-bouldin index (DBI), clustering, k-means, k-medoids, rapidminer*

Abstract

Since 2005 floods have been a problem in Rokan Hilir Regency until now, the worst floods ever occurred in the 2011-2014 year which caused many areas and residents to be directly affected by this, so the community needs help from several parties, especially the government agencies such as the social services, to help reduce the burden on society social services have one of the work programs that is to provide direct assistance both morally and logistically for flood victims in the district of Rokandownstream. For social services to assist on time, location, target, quality, quantity, and according to community needs, clustering is needed in flood-prone areas. In this study, flood-prone areas in the Rokan Hilir district were grouped. To carry out clustering in this study, the K-Means and K-Medoids algorithms will be used. After conducting experiments using RapidMiner tools on clusters K=2 to K=6, it was found that the K-Means method was more optimal compared to using the K-Medoids method on flood event data in Rokan Hilir in 2019-2022 with the optimal number of k being k = 3 with a Davies-Bouldin Index (DBI) validity value of 0.218%, while the optimal cluster K-Medoids algorithm is at K=4 with a validityvalue of 0.525%.

Keyword: flood, davies-bouldin index (DBI), clustering, k-means, k-medoids, rapidminer

1. PENDAHULUAN

Negara Indonesia merupakan negara terluas ke-14 sekaligus Negara kepulauan terluas di dunia yang letaknya berada diantara dua benua dan dua samudera yang mana merupakan pertemuan anatar lempeng tektonik Indo-Australia, Pasifik dan Eurasia. Hal ini menyebabkan banyak daerah di Indonesia rawan terhadap banjir [1][2].

Kabupaten Rokan Hilir memiliki 18 kecamatan dan 156 desa, wilayah ini merupakan daerah yang dilalui aliran sungai atau yang disebut Daerah Aliran Sungai (DAS) yaitu sungai rokan [3], Sungai Rokan Kanan juga mengalir melalui Kabupaten Rokan Hilir dan merupakan pertemuan dengan Sungai Rokan Kiri. Rokan kanan berperan penting sebagai transportasi penduduk dan sebagai sumber perekonomian masyarakat. Bergantung pada kondisi topografi dan morfologi, banjir terjadi setiap tahun di Rokan Kanan. Selain curah hujan, banjir juga disebabkan oleh faktor lain seperti topografi wilayah, struktur geologi, badan air (drainase) dan perubahan lingkungan [4].

Banjir merupakan kondisi dimana lahan atau tanah pada wilayah daratan yang biasanya kering kemudian tergenang oleh air baik pada permukiman penduduk maupun pertanian. Banjir dapat terjadi ketika Hujan yang besar menyebabkan luapan yang menyebabkan luapan sehingga tergenangnya air dalam waktu yang lama, banjir juga dapat terjadi ketika sungai tidak dapat menampung air karena volume dan debit air yang besar serta beberapa faktor lainnya [5]. Menurut Data Badan Nasional Penanggulangan Bencana (BNPB), bencana yang sering terjadi di Indonesia yaitu banjir. Tercatat setiap tahun banjir disertai longsor menjadi urutan ke-6 yang sering terjadi dengan 32 kejadian pertahunnya [6].

Sejak tahun 2005 banjir menjadi permasalahan di Kabupaten Rokan Hilir hingga saat ini, banjir terparah pernah terjadi pada tahun 2011-2014 yang menyebabkan banyak daerah dan warga secara langsung terdampak akibat hal tersebut, sehingga masyarakat sangat membutuhkan bantuan dari beberapa pihak terutama pihak instansi pemerintahan seperti dinas sosial, untuk membantu mengurangi beban masyarakat dinas sosial memiliki salah satu program kerja yaitu memberikan bantuan secara langsung baik moral maupun logistic bagi para korban banjir di kabupaten rokan hilir. Agar dinas sosial dapat memberikan bantuan yang tepat waktu, lokasi, sasaran, kualitas, kuantitas, dan sesuai kebutuhan masyarakat, dibutuhkan Clustering data untuk membantu mengelompokkan daerah yang rawan terhadap banjir yaitu pengolahan *data mining*.

Data mining adalah suatu cara yang dilakukan guna mendapatkan suatu keputusan serta mendapat kan hubungan antar data yang sebelumnya belum diketahui dengan menggunakan cara yang mudah dipahami [7][8]. Data mining memiliki beberapa bagian yaitu association, estimation, classification serta clustering. Clustering yaitu salah satu cara yang ada di data mining yang bertujuan untuk mengelompokkan suatu data (objek) kedalam suatu kelompok yang mana nantinya objek yang serupa akan digabungkan kedalam cluster yang sama sedangkan objek lainnya yang berbeda akan digabungkan kedalam cluster lainnya [9]. Sehingga memaksimalkan homogenitas (eksternal) homogenitas (internal) dalam cluster dan antar data. Terdapat beberapa algoritma clustering yang ada di data mining seperti K-Means clustering dan K-Medoids

Penelitian yang dilakukan oleh Nurdin, Fakhri dan Sri Djuniati mengenai analisis daerah rawan banjir di kabupaten Rokan Hilir dan Rokan Hulu, didapat 0,9%, kategori tidak berbahaya 56,33%, kategori berbahaya adalah 30,32% dan kategori sangat rawan bencana banjir 12,46%. Penelitian yang dilakukan oleh Devy, dkk(2022) mengenai pengelompokan daerah rawan bencana menggunakan algoritma K-Means, didapat bahwa metode K-Means dapat mempermudah dalam menganalisis daerah rawan bencana, diperoleh kluster optimal yaitu K=5 dengan nilai validitas sebesar 0.619 [10]. Penelitian yang dilakukan oleh Tri Juninda, dkk (2019) mengenai pengelompokan penyakit di Pekanbaru menggunakan algoritma K-Medoids, didapat bahwa algoritma K-Medoids dapat mengelompokkan cluster terbaik dengan penyakit dominan pada cluster tersebut, Kluster paling optimal berada pada K=4 dengan nilai validitas sebesar 0.043 [11].

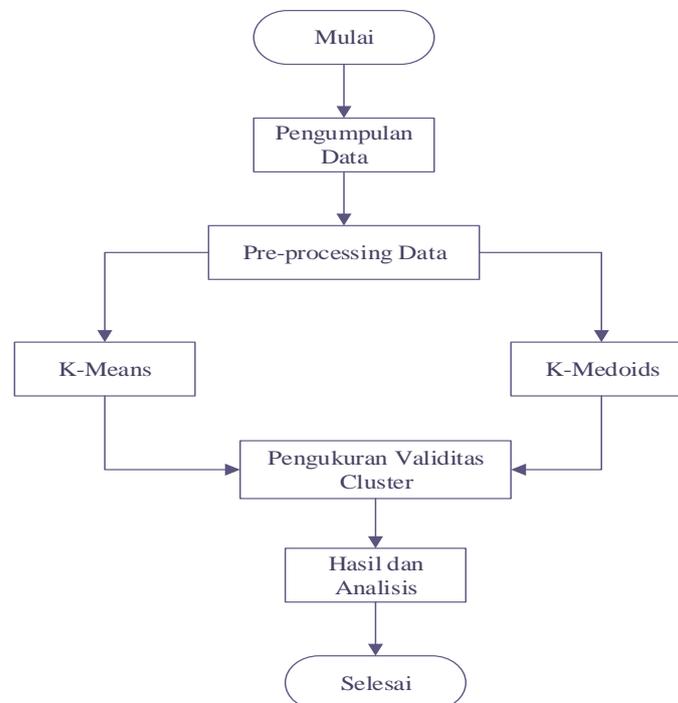
Berdasarkan penelitian yang dilakukan oleh Devy, dkk(2022) mengenai pengelompokan daerah rawan bencana menggunakan algoritma K-Means, didapat bahwa metode K-Means dapat mempermudah dalam menganalisis daerah rawan bencana, diperoleh kluster optimal yaitu K=5 dengan nilai validitas sebesar 0.619 [10]. Penelitian yang dilakukan oleh Tri Juninda, dkk (2019) mengenai pengelompokan penyakit di Pekanbaru menggunakan algoritma K-Medoids, didapat bahwa algoritma K-Medoids dapat mengelompokkan cluster terbaik dengan penyakit dominan pada cluster tersebut, Kluster paling optimal berada pada K=4 dengan nilai validitas sebesar 0.043 [11].

Penelitian lain mengenai perbandingan algoritma K-Means dan K-Medoids juga dilakukan oleh Dini, dkk (2018) mengenai pengelompokan wilayah sebaran cacat pada anak, diperoleh kesimpulan bahwa algoritma K-Medoids mampu melakukan klustering penyebaran cacat anak dengan nilai validitas sebesar 0.5009 sedangkan K-Means dengan nilai validitas sebesar 0.1443. Hal tersebut menunjukkan bahwa algoritma K-Medoids lebih optimal dibandingkan algoritma K-Means dalam pengelompokan cacat pada anak [12].

Berdasarkan permasalahan yang sudah dijelaskan pada pendahuluan maka dibuatlah suatu pengelompokan terhadap daerah rawan banjir yang ada di kabupaten Rokan Hilir. Oleh karena itu pada penelitian ini diangkatlah sebuah judul mengenai perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Daerah Rawan Banjir di Kabupaten Rokan Hilir

2. BAHAN DAN METODE PENELITIAN

Terdapat lima tahapan utama yang akan dilakukan pada penelitian ini yakni tahap pengumpulan data, dataset yang telah didapat akan melakukan proses *pre-processing* data, data yang melewati tahap perbaikan kualitas data akan diproses menggunakan kedua algoritma yaitu algoritma *K-Means* dan *K-Medoids*, setiap algoritma yang telah diproses akan dihitung validitas kluster menggunakan perhitungan DBI, dari hasil klustering dan validitas DBI akan dianalisis guna mendapatkan hasil [13]. Metodologi penelitian ini dapat dilihat pada gambar 2.

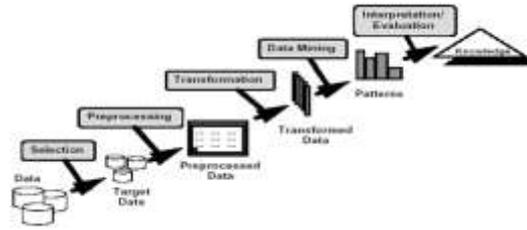


Gambar 1. Metodologi Penelitian

Pada penelitian ini dataset yang digunakan adalah data korban banjir di Kabupaten Rokan Hilir yang diperoleh langsung dari Dinas Sosial kabupaten Rokan Hilir pada bulan Januari 2019 hingga Mei 2022. Selanjutnya data akan masuk ke tahap *pre-processing* dengan melakukan cleaning data. Setelah data memiliki kualitas yang baik selanjutnya data akan diproses menggunakan kedua algoritma yakni algoritma *K-Means* dan *K-Medoids* yang diolah menggunakan *tools rapidminer*, tahap selanjutnya, proses analisis dan hasil dilakukan perhitungan *validitas* kluster dengan menggunakan *Davies Bouldin Index* (DBI), kemudian didapatkan hasil algoritma terbaik.

2.1 Data Mining

Data Mining merupakan penggalian informasi yang untuk menemukan pola dataset yang unik dan bermanfaat dengan cara memproses data tersebut menggunakan algoritma atau metode yang sesuai dengan tujuan pada proses KDD (*Knowledge Discovery in Database*) [14]. Adapun tujuan data mining yakni untuk menemukan pola dataset yang besar yang digunakan untuk membantu dalam pengambilan keputusan nantinya. Berikut merupakan gambar tahapan data mining untuk mencari pola.



Gambar 2. Proses Data Mining

2.2 Clustering

Clustering merupakan pengelompokan suatu kelas yang memiliki kesamaan bentuk, sedangkan perbedaan dengan klasifikasi yakni tidak adanya variabel target untuk pengelompokan pada *clustering*. Clustering dilakukan sebagai tahap awal dalam proses data mining saat melakukan tahapan analisis. Adapun banyak algoritma untuk proses klasifikasi seperti algoritma Fuzzy C-Means, K-Means, K-Medoids, Clarans dll. Masing-masing algoritma memiliki keunggulan maupun kelemahan jika dibandingkan dengan algoritma lain, akan tetapi masing-masing metode data mining memiliki kesamaan prinsip yakni mengelompokkan data dengan mengukur jarak kesamaan dan mencari karakteristik yang sesuai dengan data dalam satu kelompok. [16]

2.3 K-Means

K-Means merupakan metode data mining untuk mengelompokkan dataset melalui titik pusat (*Centroid*) dengan data yang terdekat [17]. *K-Means* bertujuan untuk memaksimalkan kemiripan dalam suatu kluster dan meminimalkan kesamaan data antar kluster yang berbeda. Parameter yang digunakan untuk memaksimalkan kemiripan data pada cluster adalah fungsi jarak sehingga kesamaan data didasarkan pada jarak terdekat dari *centroid* [18]. Ada beberapa tahapan dalam melakukan perhitungan menggunakan algoritma *K-Means* yaitu:

1. Menentukan nilai *K* untuk titik awal kluster secara acak
2. Masing-masing data kemudian akan dibagi ke setiap kluster antar titik pusat menggunakan jarak *Euclidean*. Rumus *Euclidean* dapat dilihat pada persamaan (1) [9]:

$$d(x,y) = \|x-y\|^2 = \dots\dots\dots(1)$$

dengan :

- d= jarak data ke pusat cluster
- x= data ke pada atribut
- y= titik pusat ke, pada atribut

3. masing-masing pusat kluster akan dilakukan perhitungan ulang dari nilai rata-rata didalam kluster yang telah didapatkan.
4. selanjutnya tahap 2 dan 3 akan dilakukan perulangan sampai nantinya ditemukan kelompok yang sesuai dan tidak adanya perubahan antar anggota dalam cluster dengan percobaan sebelumnya.

2.4 K-Medoids

Algoritma *K-Medoids* atau *PAM (Partitioning Around Medoids)* adalah metode data mining yang diwakilkan oleh kluster yakni *medoid*. Perbedaan dengan algoritma *K-Means* yakni memerlukan nilai rata-rata yang dijadikan pusat kluster. Berikut merupakan tahapan yang dapat dilakukan pada Algoritma *K-Medoids*.

Langkah-langkah penyelesaian *K-Medoids* adalah sebagai berikut:

1. Menentukan jumlah *K* sebagai pusat kluster
2. Setiap data dimasukkan kedalam kluster terdekat dengan mengukur jarak *Euclidian Distance* dengan rumus persamaan (1).
3. Pilih record data pada setiap kluster secara acak untuk dijadikan *medoid* yang baru
4. Objek yang termasuk kedalam kluster masing-masing akan dilakukan perhitungan dengan bakal *medoid* baru.
5. Total simpangan akan dihitung dengan total jarak baru ke total jarak lama, jika didapat $S < 0$, maka cari data baru untuk membuat *k* objek yang baru sebagai *medoid*.
6. Lakukan perulangan pada langkah 3 hingga *medoid* tidak berubah, sehingga didapat anggota pada masing-masing kluster [11].

2.5 Davies-Bouldin Index (DBI)

Davies Bouldin Index (DBI) adalah salah satu cara yang digunakan untuk menghitung validitas kluster. Perhitungan nilai DBI dilakukan dengan nilai *separasi* dan *kohesi*. Kohesi yakni jumlah kedekatan hubungan data dalam sebuah kluster, sedangkan *sparasi* yakni jarak perbedaan jarak tiap kluster dengan kluster lainnya. *Cluster* yang paling optimal adalah kluster dengan nilai kohesi besar dan nilai sparasi yang rendah [6]. Apabila nilai *Davies Bouldin Index* (DBI) mendekati 0 maka semakin baik cluster yang diperoleh. Itu artinya semakin rendah nilai DBI maka semakin bagus validitas klusternya [19].

3. ANALISA DAN HASIL

3.1 Pengumpulan dan Preprocessing Data

Data yang digunakan pada penelitian ini adalah data Banjir di Kabupaten Rokan Hilir yang diperoleh dari Dinas Sosial pada Januari 2019 Hingga Mei 2022. dari 4 atribut yang ada pada data korban banjir di kabupaten rokan hilir, diambil 3 aribut yaitu kecamatan, desa dan jumlah jiwa yang terdampak dengan jumlah kejadian terkonfirmasi sebanyak 84 data desa serta 66.670 Jumlah korban. Data hasil pengelompokan perdesa dan cleaning adalah data yang akan diproses pada penelitian ini yang terdapat pada tabel 1.

Tabel 1. Cleaning Data

N0	DESA	KECAMATAN	JUMLAH JIWA
1	Pujud	Pujud	3049
2	Pujud	Pujud	68
3	Rantau Kopar	Rantau Kopar	3521
4	Bangko Bagan	Bangko Bagan	1736
5	Sinembah	Sinembah	56
6	Balai Jaya	Balai Jaya	116
7	Bangko Pusako	Bangko Pusako	236
.....
53	Pekaitan	Pekaitan	84

Setelah tahap *cleaning* data dilakukan, selanjutnya data akan masuk pada tahap *transformasi* yakni dengan mengubah data dari angka (*nominal*) kedalam bentuk (*numerik*), ini dilakukan agar proses perhitungan dapat dilakukan lebih mudah nantinya. Penetapan perubahan data kedalam bentuk *numerik* dapat dilihat pada tabel 2.

Tabel 2. Transformasi Kecamatan

Kecamatan	Transformasi
Bagan Sinembah	1
Balai Jaya	2
Bangko	3
Bangko Pusako	4
Kubu Babussalam	5
Pasir Limau Kapas	6
Pekaitan	7
Pujud	8
Rantau Kopar	9
Rimba Melintang	10
Simpang Kanan	11
Tanah Putih	12
Tanah Putih Tanjung	
Melawan	13
Tanjung Medan	14

Berdasar tabel 2 didapat data transformasi secara keseluruhan, yang dapat dilihat pada tabel 3:

Tabel 3. Transformasi Keseluruhan

NO	DESA	KECAMATAN	JUMLAH KORBAN
1	Air Hitam Babussalam	1	3049
2	Rokan Bagan	2	68
3	Cempedak	3	3521
4	Bagan Hulu	4	1736
5	Bagan Sapta Permai	5	56
6	Balai Jaya Kota	5	116
7	Bangko Kanan	5	236
...
53	Pekaitan	15	84

3.2 Proses Data

Data yang telah melalui tahap *pre-poccesing*, kemudian data akan diproses menggunakan *tools rapidminer* yang akan digunakan untuk mendapatkan validitas dengan perhitungan *DaviesBouldin Index* (DBI) sebagai pedoman untuk mendapatkan kluster yang paling *optimal*. Pengujian dilakukan pada cluster $k=2$ sampai dengan $k=6$.

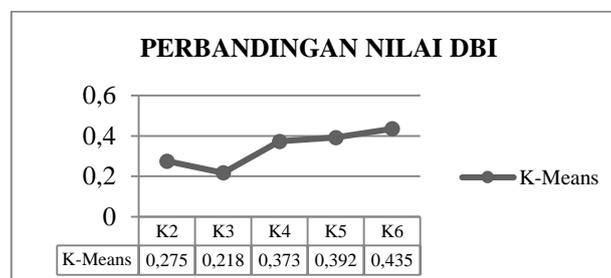
3.3 K-Means

Proses Clustering pada penelitian ini menggunakan algoritma K-Means. Berdasarkan penelitian yang dilakukan oleh Mufida, dkk (2021), pada penelitiannya menggunakan percobaan cluster 2, 3, 4, 5 dan 6. Hasil cluster tersebut dapat dilihat pada tabel 4.

Tabel 4. Hasil Klasterng K-Means

Percobaan	Klaster					
	1	2	3	4	5	6
K=2	41	12	-	-	-	-
K=3	21	3	29	-	-	-
K=4	8	29	2	14	-	-
K=5	7	2	5	15	24	-
K=6	5	2	24	10	5	17

Setelah dilakukan proses pengelompokan menggunakan algoritma k-means kemudian dilakukan validitas cluster dengan teknik DBI. Nilai DBI berdasarkan hasil klastering data menggunakan algoritma K-Means tertera pada gambar 3.



Gambar 3. Nilai DBI K-Means

Dapat dilihat pada Gambar 3 bahwa Berdasarkan nilai DBI terbaik, cluster terbaik pada algoritma k-means terletak pada percobaan dengan $k=3$ dengan nilai DBI yang diperoleh sebesar 0,218 yang membagi data menjadi 3 cluster dengan 21 anggota pada cluster 1, 3 anggota pada cluster 2, dan 29 anggota pada cluster 3.

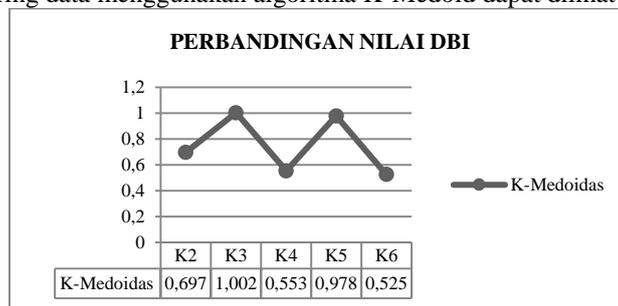
3.4 K-Medoids

Proses Clustering pada penelitian ini juga menggunakan algoritma K-Medoid dengan menggunakan percobaan cluster yang sama dengan algoritma sebelumnya, sehingga juga menggunakan percobaan cluster 2, 3, 4, 5, dan 6. Hasil cluster tersebut dapat dilihat pada tabel 5

Tabel 5. Hasil Klastering *K-Medoids*

Percobaan	Klaster					
	1	2	3	4	5	6
K=2	26	27	-	-	-	-
K=3	11	15	27	-	-	-
K=4	9	2	27	15	-	-
K=5	12	21	4	9	7	-
K=6	2	9	15	11	4	12

Kemudian juga dilakukan validitas cluster dengan menggunakan Davies Bouldin Index (DBI). Nilai DBI berdasarkan hasil Clustering data menggunakan algoritma K-Medoid dapat dilihat pada gambar 4.

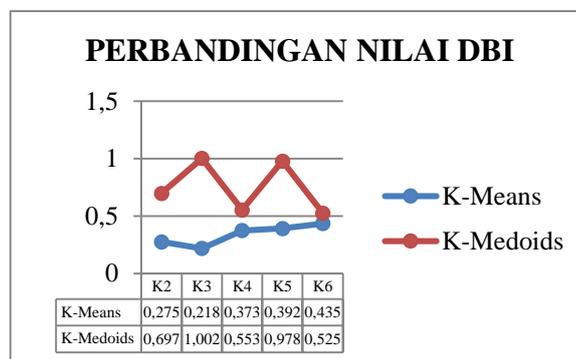


Gambar 4. Nilai DBI K-Means

Dapat dilihat pada Gambar 4.2 Berdasarkan nilai DBI terbaik, cluster terbaik pada algoritma k-medoid terletak pada percobaan dengan k=6 dengan nilai DBI yang diperoleh adalah 0.525. Pada percobaan nilai k=6 membagi data menjadi 6 cluster dengan 2 anggota pada cluster 1, 9 anggota pada cluster 2, 15 anggota pada cluster 3, 11 anggota pada cluster 4, 4 anggota pada cluster 5 dan 12 anggota pada cluster 6.

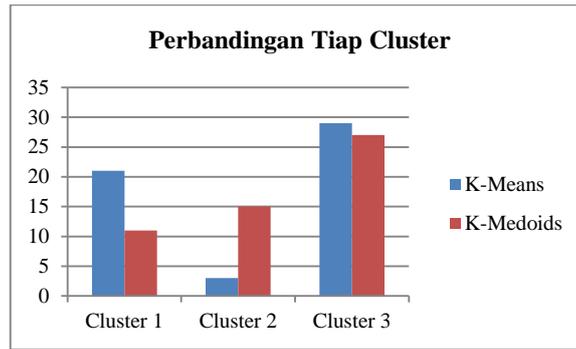
4. Perbandingan Validitas Cluster

Setelah melakukan percobaan pengelompokan dengan 2 Algoritma yaitu K-Means dan K-Medoids, langkah selanjutnya menentukan klaster terbaik dengan cara membandingkan nilai validitas dengan DBI pada setiap percobaan algoritma. Dari percobaan yang telah dilakukan hasilnya dapat dilihat pada gambar 5 sebagai berikut :



Gambar 5. Perbandingan Nilai DBI

Berdasarkan perbandingan nilai *Davies-Bouldin Index* diatas, K-Means memperoleh nilai terkecil pada k=3 dengan nilai *validitas* DBI sebesar 0.218, sedangkan *K-Medoids* DBI terkecil terletak pada k=6 dengan nilai sebesar 0,525. Dengan demikian diketahui bahwa nilai *K-Means* dengan k=3 lebih *optimal* dibandingkan dengan nilai *K-Medoids* dengan k=6, maka pengelompokan akan dilakukan dengan algoritma *K-Means* pada data bencana banjir dengan jumlah paling optimal adalah k=3 dapat ditunjukkan pada gambar 6.



Gambar 6. Perbandingan Tiap Cluster

Detail dari hasil pengelompokan yang dilakukan dengan algoritma *K-Means* pada data banjir dengan nilai $k=3$ dapat dilihat pada tabel 6.

Tabel 6. Hasil Pengelompokan Menggunakan *K-Means*

Cluster	Jumlah Data	Anggota Kelompok
1	21	Air Hitam, Bagan Hulu, Bangko Makmur, Bangko Mukti, Karya Mukti, Karya Mulyo Sari, Kubu I, Labuhan Papan, Putat, Rantau Bais, Rokan Baru, Sedinginan, Sekapas, Sintong Pusako, Sungai Besar, Sungai Manasib, Sungai Panji-Panji, Sungai Rangau, Ujung Tanjung, Ulak Kembahang, Sei Rangau
2	3	Bagan Cempedak, Rantau Kopar, Melayu Besar
3	29	Babussalam Rokan, Bagan Sapta Permai, Balai Jaya Kota, Bangko Kanan, Bhayangkara Jaya, Bukit Selamat, Jumrah, Kasang Bangsawan, Kelurahan Pujud Selatan, Mesah, Pematang Botam, Pujud Selatan, Rokan Baru Pesisir, RTPK Hilir, Sekeladi, Sekeadi Hilir, Siarang-Arang, Siarang-Arang Rokan, Sintong, Sintong Bakti, Sungai Daun, Sungai Pinang, Sungai Tapah, Teluk Bano I, Teluk Bano II, Teluk Berembun, Mumugo, Pdamaran, Pekaitan.

Pada tabel 6 diketahui bahwa klaster dengan $k=3$ adalah klaster dengan jumlah daerah terbanyak. Namun, jumlah korban paling tinggi terdapat pada klaster $k=2$ dengan jumlah 3 daerah dan jumlah korban terdampak sebanyak 13.463 korban, ketiga daerah tersebut yaitu Bagan Cempedak, Rantau Kopar dan Melayu Besar. Daerah tersebut berada pada kecamatan Tanah Putih tanjung Melawan dan Rantau Kopar, menurut profil dari kedua kecamatan tersebut kecamatan Tanah Putih tanjung Melawan dan Rantau Kopar merupakan daerah yang termasuk kedalam daerah yang dilalui oleh aliran sungai atau yang disebut DAS, memiliki struktur tanah gambut dan tanahnya merupakan daerah dengan dataran rendah sehingga menjadi faktor mengapa banjir sering kali terjadi di daerah tersebut.

4. KESIMPULAN

Berdasarkan hasil dan analisis, didapat pengelompokan daerah rawan banjir di Rokan Hilir pada tahun 2019-2022 dapat dilakukan dengan algoritma *K-Means* maupun *K-Medoids*, akan tetapi nilai validitas DBI klaster yang paling optimal terdapat pada algoritma *K-Means* yakni 0.218 dengan klaster paling optimal adalah $k=3$, sedangkan algoritma *K-Medoids* klaster optimal terdapat pada $K=6$ dengan nilai validitas sebesar 0.525. Perolehan cluster dominan, bahwa klaster 3 merupakan klaster dengan jumlah daerah paling banyak. Sedangkan jumlah kejadian terbanyak terletak pada kluster 2 dengan jumlah daerah adalah 3 daerah dan jumlah jumlah jiwa yang terdampak sebanyak 13.463 korban, ketiga daerah tersebut yaitu Bagan Cempedak, Rantau Kopar dan Melayu Besar. Daerah tersebut berada pada kecamatan Tanah Putih tanjung Melawan dan Rantau Kopar, menurut profil dari kedua kecamatan tersebut kecamatan Tanah Putih tanjung Melawan dan Rantau Kopar merupakan daerah yang termasuk kedalam daerah yang dilalui oleh aliran sungai atau yang disebut DAS, memiliki struktur tanah gambut dan tanahnya merupakan daerah dengan dataran rendah sehingga menjadi faktor mengapa banjir sering kali terjadi di daerah tersebut.

REFERENSI

- [1] A. Febriansyah *et al.*, "Penerapan Machine Learning Dalam Mitigasi Banjir Menggunakan Data Mining," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 215–218, 2020, doi: 10.32672/jnkti.v3i3.2427.
- [2] M. I. Ramadhan, "Penerapan Data Mining untuk Analisis Data Bencana Milik BNPB Menggunakan Algoritma *K-Means* dan Linear Regression," *J. Inform. dan Komput.*, vol. 22, no. 1, pp. 57–65, 2017.

- [3] Nurdin, Fakhri, and S. Djuniati, "Analisis kawasan rawan banjir sungai rokan dalam kabupaten rokan hulu dan kabupaten rokan hilir," pp. 168–175, 2018.
- [4] R. Novita, H. Kadir, and E. Eriyati, "Nilai Kerugian Masyarakat Akibat Banjir di Kecamatan Pujud Kabupaten Rokan Hilir." Riau University, 2014.
- [5] A. I. Suryani and N. Afriansih, "Pengetahuan Mitigasi Masyarakat Tentang Bencana Banjir Di Nagari Aur Begalung Talaok Kaecamatan Bayang Kabupaten Pesisir Selatan," *J. Spasial*, vol. 3, no. 6, p. 87, 2019.
- [6] S. Dwiasnati and Y. Devianto, "Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir," *Pros. SISFOTEK*, pp. 202–207, 2021, [Online]. Available: <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283>
- [7] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.
- [8] A. Qurotul, E. Tasia, N. Nazira, P. F. Pratama, M. R. Anugrah, and J. Adhiva, "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," vol. 4, no. September, pp. 72–76, 2022, doi: 10.30865/json.v4i1.4781.
- [9] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [10] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," vol. 9, no. 3, pp. 749–753, 2022, doi: 10.30865/jurikom.v9i3.4326.
- [11] T. Juninda and E. Andri, "Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau," no. November, pp. 42–49, 2019.
- [12] D. Marlina, N. F. Putri, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," vol. 4, no. 2, pp. 64–71, 2018.
- [13] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. Fitri Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022.
- [14] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.
- [15] D. F. Pramesti, M. T. Furqon, and C. Dewi, "GPU accelerated Krylov subspace methods for computational electromagnetics," *Proc. 38th Eur. Microw. Conf. EuMC 2008*, vol. 1, no. 9, pp. 1312–1314, 2008, doi: 10.1109/EUMC.2008.4751704.
- [16] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [17] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteraan untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [18] I. N. M. Adiputra, "Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, p. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [19] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.