Institute of Research and Publication Indonesia
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 1 Iss. 1 April 2021, pp: 34-40
P-ISSN: 2797-2313 E-ISSN: 2775-8575

Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province

Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat

Mufidah Herviany[1], Saleha Putri Delima[2] , Triyana Nurhidayah[3], Kasini[4]

[1,2,3]Program Studi Sistem Informasi, Fakultas Sains dan Teknologi UIN Sultan Syarif Kasim Riau
Jl. HR Soebrantas KM. 18 No. 155 Panam Pekanbaru, Indonesia
[4]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi
Universitas Pahlawan Tuanku Tambusai Bangkinang, Riau, Indonesia

E-mail: [1]11850320470@students.uin-suska.ac.id, [2]11850320455@students.uin-suska.ac.id,
[3]11850320433@students.uin-suska.ac.id, [4]kasini@gmail.gom

Received Februari 03th 2021; Revised February 18th 2021; Accepted February 27th 2021
Corresponding Author: Mufidah Herviany

Abstract

A natural disaster is an incident that cannot be avoided. However, the effects of disasters can be reduced by identifying the triggers for disasters as well as reviewing disaster events that have occurred through analysis of existing disaster data. Indonesia often experiences disasters caused by natural damage caused by human actions such as floods and landslides. Based on data from the West Java Open Data in the 2019 period, West Java province experienced 609 landslide events. The Regional Disaster Management Agency (BPBD) has not been able to optimize services for disaster victims, for example the length of time to arrive for assistance due to limited equipment and food in the disaster area. Meanwhile, the existence of disaster risk mapping is very important in structuring targeted and precise disaster management. So it is necessary to process data to determine district / city areas where landslides often occur. Researchers used data processing using the K-Means and K-Medoids algorithm comparison method. The method obtained from grouping with the K-Means method is more optimal than using the K-Medoids method in West Java Province landslide incidence data in 2019 with the most optimal number of k is k = 6. The acquisition of dominant clusters shows that cluster 2 is a cluster with the largest number of regions. And the largest number of incidents is located in cluster 5 with the number of regions is 4 regions and the number of incidents is 106 events.

Keywords: *Clustering, Davies-Bouldin Index, K-Means, K-Medoids*, *Landslides.*

Abstrak

Bencana alam ialah insiden yang tidak bisa dihindari. Tetapi akibat dari bencana bisa dikurangi dengan mengidintifikasi pemicu terjadinya bencana serta mengkaji peristiwa bencana yang sudah pernah terjadi melalui analisa data bencana yang ada. Indonesia sering mengalami bencana yang disebabkan oleh kerusakan alam akibat perbuatan manusia seperti bencana banjir dan tanah longsor. Berdasarkan data dari Jabar Open Data pada periode 2019, provinsi jawa barat mengalami 609 peristiwa tanah longsor. Badan Penanggulangan Bencana Daerah (BPBD) belum dapat mengoptimalkan pelayanan terhadap korban bencana, misalnya lamanya datang bantuan karena terbatasnya peralatan dan makanan pada daerah bencana. Sedangkan dengan adanya pemetaan resiko bencana menjadi sangat penting dalam penataan penanggulangan bencana yang terarah dan tepat. Maka diperlukannya pengolahan data untuk mengetahui daerah kabupaten/kota yang sering terjadi bencana tanah longsor. Peneliti menggunakan pengolahan data dengan metode perbandingan algoritma *K-Means* dan *K-Medoids*. Metode yang didapatkan dari pengelompokkan dengan *method K-Means* lebih optimal daripada mengguanakan *method K-Medoids* pada data kejadian tanah longsor Provinsi Jawa Barat pada tahun 2019 dengan jumlah k paling optimal adalah k = 6. Perolehan *cluster* dominan, menunjukkan bahwa kluster 2 merupakan kluster dengan jumlah daerah paling banyak. Dan jumlah kejadian terbanyak terletak pada kluster 5 dengan jumlah 4 daerah dan jumlah kejadian sebanyak 106 kejadian.

Kata Kunci: Clustering, Davies-Bouldin Index, K-Means, K-Medoids, Tanah Longsor.

1. PENDAHULUAN

Bencana alam ialah suatu insiden yang tidak bisa dihindari. Akan tetapi akibat dari bencana bisa dikurangi dengan mengidentifikasi pemicu terjadinya bencana serta mengkaji peristiwa bencana yang sudah pernah terjadi melalui analisa data bencana yang ada. Menurut Undang-undang Nomor 24 Tahun 2007, bencana adalah serangkaian peristiwa yang dapat mengancam serta terganggunya kehidupan masyarakat disebabkan karena faktor alam ataupun faktor nonalam serta faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, rusaknya lingkungan, kerugian harta benda, serta dampak pada psikologis [1].

Di Indonesia terdapat banyak daerah yang penduduknya sangat tinggi rawan akan bencana alam. Faktor geologi merupakan salah satu faktor yang mengakibatkan bencana alam, selain faktor geolofi tersebut Indonesia juga sering mengalami bencana yang disebabkan oleh perbuatan manusia yang merusak alam, seperti banjir dan tanah longsor [2]. Berdasarkan data-data dari Jabar Open Data pada periode 2019, provinsi jawa barat mengalami 609 peristiwa bencana tanah longsor.

Bencana longsor adalah satu dari beberapa jenis bencana yang terjadi di Indonesia. Akibatnya, terjadi pendangkalan, rusaknya lahan pertanian dan pemukian serta terganggunya jalur lalu lintas. Dalam hal ini Badan Penanggulangan Bencana Daerah (BPBD) belum dapat mengoptimalkan pelayanan terhadap korban bencana, misalnya lamanya datang bantuan karena terbatasnya peralatan dan makanan pada daerah bencana. Sedangkan dengan adanya pemetaan resiko bencana menjadi sangat penting dalam penataan penanggulangan bencana yang terarah dan tepat.

Berdasarkan uraian permasalahan tersebut hingga diperlukan beberapa kelompok daerah kabupaten/kota di provinsi Jawa Barat bersumber pada tingkat banyaknya peristiwa alam tanah longsor pada tahun tersbut, maka dari akan terbentuk pemetaan daerah-daerah rawan bencana alam tanah longsor semua kabupaten/kota pada provinsi Jawa Barat dapat dijadikan acuan oleh BPBD ataupun pemerintahan yang terkait persiapan prabencana agar dapat menanggulangi akibat dari kejadian bencana tanah longsor pada manusia dan hata benda.

Penelitian terkait dengan perbandingan *algorithm K-Means* dan *K-Medoid* dalam menentukan hasil *cluster*, pernah dilakukan oleh beberapa peneliti sebelumnya. Pada tahun 2019 [3] melakukan penelitian "Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau" didapatkan kesimpulan bahwa dalam perbandingan dua algoritma ini tidak menunjukkan perbedaan yang terlalu signifikan dalam pengelompokkan datanya. Pada tahun 2012 [4] melakukan penelitian "Efficiency of *k-Means* and *K-Medoids* Algorithms for *Clustering* Arbitrary Data Points " didapatkan kesimpulan bahwa waktu komputasi algoritma kMeans lebih kecil daripada algoritma *K-Medoids* untuk aplikasi yang dipilih. Dan juga pada penelitian lainnya [5], [6], [7].

Berdasarkan latar belakang masalah tersebut, dalam riset ini dilakukan perbandingan *algorithm K-Means* dan *K-Medoids* menentukan pengelompokkan (*clustering*) data bencana alam tanah longsor yang sering terjadi pada daerah kabupaten/kota provinsi Jawa barat.

2. METODOLOGI PENELITIAN

Berlandakskan pada pendahuluan diatas, dalam penelitian ini terdapat beberapa tahapan, tahapan metodologi penelitian yang digunakan dapat dilihat pada gambar 1.

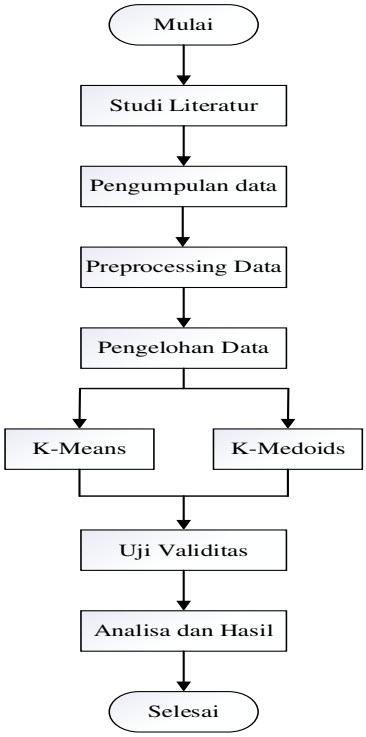
Riset ini didahului dengan melakukan studi literatur, pengumpulan data dan setelahnya melakukan *preprocessing* data yaitu *cleaning* data untuk nantinya diolah dengan melakukan perbandingan *algorithm K-Means* dan *K-Medoids*. Pada tahap perhitungan menggunakan *tools Rapidminer*. Jika hasil dari *cluster* telah didapatkan maka akan dilakukan analisis pada *cluster* dan data-data yang ada didalamnya sehingga akan didapatkan hasil dan kesimpulan.

2.1 *Data Mining*

Data mining merupakan analisis pada suatu data yang bertujuan untuk mendapatkan keterkaitan yang jelas dan mendapatkan kesimpulan yang belum diketahui sebelumnya dengan metode terbaru serta bermanfaat untuk pemilik data tersebut. Data mining merupan metode yang ditujukan untuk mengekstrak sebuah informasi prediktif tersembunyi di *database,* ini merupakan teknologi yang potensial bagi perusahaan yang sangat potensial dalam memberdayakan *data warehouse* [8].

Data Mining (DM) adalah ekstraksi informasi dari sejumlah besar data untuk melihat yang tersembunyi pengetahuan dan memfasilitasi penggunaannya secara real time aplikasi. DM memiliki berbagai macam algoritma untuk analisis data. Beberapa teknik DM utama digunakan untuk analisis adalah *Clustering*, Association, Klasifikasi dan lain-lain. *Cluster* adalah efektif teknik untuk analisis data eksplorasi, dan memiliki menemukan aplikasi di berbagai bidang. Paling metode *clustering* yang ada dapat dikategorikan menjadi tiga: partisi, hierarki, berbasis grid dan metode berbasis model. Pengelompokan

berbasis partisi menghasilkan partisi data sehingga objek dalam file *cluster* lebih mirip satu sama lain daripada yang seharusnya objek di *cluster* lain.



Gambar 1. Metodologi Penelitian

2.2 *Clustering*

Clustering yaitu proses pembagian data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan dari datanya terhadap suatu kelompok lebih besar dibandingkan kesamaan dari data tersebut dengan data dari kelompok yang lain. [9].

Clustering bisa dianggap yang paling penting masalah belajar tanpa pengawasan; jadi, seperti yang lainnya masalah semacam ini, ini berkaitan dengan menemukan struktur dikumpulan data tidak berlabel (Jain dan Dubes, 1988;Jain et al., 1999). Definisi *clustering* yang longgar bisa menjadi "proses pengorganisasian objek menjadi kelompok yang anggota serupa dalam beberapa hal ". Sebuah *cluster* adalah oleh karena itu kumpulan objek yang "serupa" di antara mereka dan "berbeda" dengan objekmilik *cluster* lain [10].

2.3 *K-Means*

Algoritma *K-means* merupakan salah satu algoritma *clustering*(pengelompokan). Kmeans *clustering* merupakan metode *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster* [11]. *K-means* mencoba meminimalkan perbedaan kesalahan kuadrat antara rata-rata *cluster* dan titik data di *cluster* itu. Misalkan kita punya beberapa titik data n-dimensi yang ingin dikelompokkan pengguna dalam k jumlah *cluster* dengan μk sebagai mean dari *cluster* tersebut, maka *k-means* direpresentasikan sebagai [12]:

$$\left[\sum_{k=1}^{k}\sum_{x\in c_k}||x_i - u_k||^2\right] \quad (1)$$

Dimana xi adalah himpunan titik data dengan i = 1, 2, 3, .., n, menjadi dikelompokkan dalam sebuah *cluster* dari satu set *cluster* yang diberikan sebagai ck dengan k = 1, 2, 3,. . . ., k. Untuk mengurangi kesalahan kuadrat, *k-means* mengalokasikan pola ke kluster k yang awalnya dipartisi. Dalam hal keanggotaan *cluster* yang tidak menentukan, *k-means* terus ulangi langkah-langkah berikut :

1) Tetapkan setiap pola ke *cluster* terdekat dan buat partisi baru
2) Hitung rata-rata *cluster* baru.

2.4 *K-Medoids*

Algoritma *K-Medoids* berperan untuk menemukan *Medoids* dalam suatu *cluster* yang menjadi titik pusat *cluster*. KMedoids lebihh kuat daripada *K-Means* seperti, pada *K-Medoids* kami mendapatkan k selaku

obyek representatife diminimalkan hasil ketidaksamaan obyek data, sedangkan *K-Means* menggunakan hasil kuadrat jarak *Euclidean* pada data benda. Dan metrik jarak ini mengurangi data yang noise dan outliers[13]. Ini adalah teknik berbasis objek yang representatif. Dalam metode ini kita memilih objek sebenarnya untuk merepresentasikan *cluster* daripada mengambil nilai rata-rata objek dalam *cluster* sebagai titik referensi. PAM (metode partisi sekitar) adalah salah satu algoritma *K-Medoids* pertama. Strategi dasar dari algoritma ini adalah sebagai berikut :

1) Temukan objek representatif untuk setiap *cluster*.
2) Selanjutnya semua obyek yang tersisa akan di kelompokkan bersama obyek represntatif paling mirip.
3) Lalu secara berulang menggantikan salah satu medoid dengan non-medoid selama "kualitas" pengelompokan diterapkan.

Algoritma *K-Medoids* secara komputasi lebih sulit daripada KMeans karena menghitung medoid menggunakan frekuensi kejadian. *K-Medoids* memiliki potensi penting karakteristik pusat mana yang berada di antara data tunjuk sendiri. Algoritme baru diusulkan untuk pengelompokan KMedoids yang berjalan seperti *K-means* algoritma dan menguji beberapa metode untuk memilih awal medoid [14].

2.5 *Davies-Bouldin Index (DBI)*

Metode yang diperkenalkan oleh David L.Davies dan Donald W.Bouldin dan penamaan metode ini menggunakan nama dengan keduanya yakni Davies-Bouldin (DBI) digunakan untuk mengevaluasi *cluster*. Evaluasi dengan Davies-Bouldin Index memiliki skema evaluasi *cluster* internal, dimana hasil *cluster* baik atau tidak dilihat dari kuantitas dan kedekatan antar hasil *cluster*. Davies-Bouldin Index merupakan salah satu *method* yang dipakai untuk mengukur kevalidan antar hasil dari *cluster* suatu *method* pengelompokan, cohesion diartikan menjadi penjumlahan relasi data dengan titik pusat *cluster* yang diikuti *cluster*. Sedangkann pemisahannya berdasarkan jarak antara antar titik pusat *cluster* dengan *cluster*. Pengukuran menggunakan Davies-Bouldin Indeks ini memaksimalkan jarak antar *cluster* antara *cluster* Ci dan Cj dan pada saat yang sama mencoba meminimalkan jarak antar titik dalam suatu *cluster*. Jika jarak antar *cluster* maksimal, artinya kesamaan karakteristik antar *cluster* kecil sehingga perbedaan antar *cluster* tampak lebih jelas. Jika jarak intra *cluster* minimum berarti setiap objek dalam *cluster* memiliki tingkat kesamaan karakteristik yang tinggi [15].

3. HASIL DAN ANALISIS

3.1 Pengumpulan dan *Preprocessing* Data.

Pada penelitian ini data yang digunakan adalah data tanah longsor pada Provinsi Jawa Barat pada tahun 2019 yang diperoleh dari website resmi Open Data Provinsi Jawa Barat yaitu https://data.jabarprov.go.id/. Berdasarkan jumlah kejadian tanah longsor yang telah terjadi di Provinsi Jawa Barat pada tahun 2018 adalah sebanyak 609 kejadian. Data tersebut akan diproses untuk mengetahui hasil pengelompokan daerah rawan tanah longsor pada Provinsi Jawa Barat. Pengelompokan data ini dapat menjadi informasi baru bagi BPBD Provinsi Jawa Barat dan berguna untuk memutuskan sebuah kebijakan kedepannya.

Dalam melakukan *clustering*. sebanyak 5 atribut yang terdapat pada data jumlah kejadian tanah longsor provinsi jawa barat, diambil 2 atribut yang akan digunakan dalam perhitungan, yakni nama kabupaten dan jumlah. Selanjutnya dilakukan *clenaning* data untuk mengurangi *noise* yang akan berpengaruh pada hasil pengolahan data. Prosedur yang dilakukan pada tahapan ini adalah menghapus data yang kosong sehingga data yang tersisa berjumlah 24 data. Berikut hasil dari cleaning data dapat dilihat pada tabel 1.

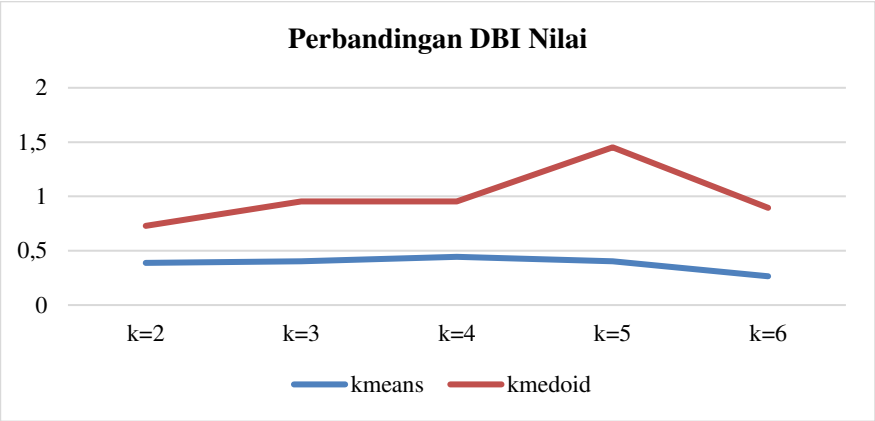
Tabel 1. Data Cleaning

Kabupaten Kota	Jumlah
Kabupaten Bandung	20
Kabupaten Bandung Barat	26
Kabupaten Bogor	68
Kabupaten Ciamis	32
Kabupaten Cianjur	7
Kabupaten Cirebon	2
Kabupaten Garut	39
Kabupaten Karawang	1
Kabupaten Kuningan	34
Kabupaten Majalengka	37
Kabupaten Pangandaran	2
Kabupaten Purwakarta	10
Kabupaten Subang	11

Kabupaten Kota	Jumlah
Kabupaten Sukabumi	100
Kabupaten Sumedang	50
Kabupaten Tasikmalaya	23
Kota Bandung	6
Kota Banjar	1
Kota Bekasi	1
Kota Bogor	112
Kota Cimahi	6
Kota Depok	1
Kota Sukabumi	2
Kota Tasikmalaya	18

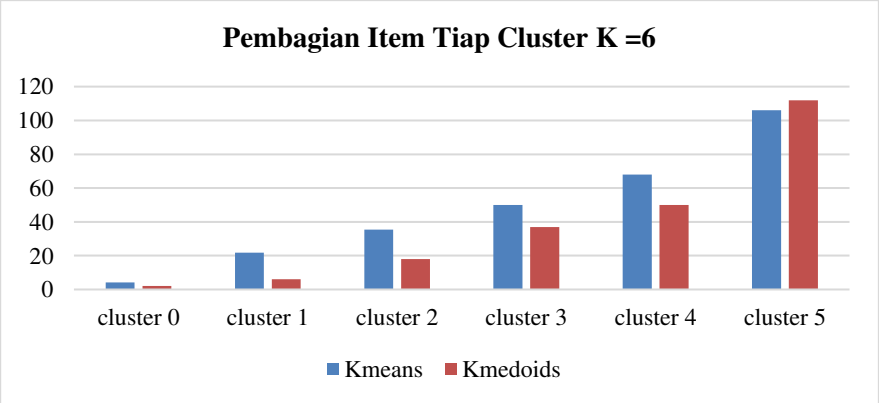
3.1 Proses Data

Setelah melakukan tahapan cleaning data, proses selanjutnya adalah menentukan jumlah *cluster* yang paling optimal dengan menggunakan Davies-Bouldin (DBI). Untuk mengetahi *clustering* terbaik pada DBI adalah dengan mengetahui nilai DBI terkecil dari pilihan yang ada. Pengujian dilakukan pada *cluster* k=2 sampai dengan k=6. Berikut ini adalah hasil pengolahan data yang tunjukkan pada gambar 2.



Gambar 2. Perbandingan Nilai DBI

Berdasarkan perbandingan nilai Davies-Bouldin Index diatas, nilai *K-Means* terkecil terletak pada k = 6 dengan nilai DBI sebesar 0,265, sedangkan nilai DBI terkecil untuk *K-Medoids* terletak pada k=2 dengan nilai DBI 0,342. Karena nilai DBIpada *cluster K-Means* (k=6) lebih kecil dari pada nilai DBI pada kluster K-Medoid (k=2), maka pengelompokkan menggunakan metode *K-Means* lebih optimal dibandingan metode *K-Medoids* pada data kejadian tanah longsor Provinsi Jawa Barat pada tahun 2019 dengan jumlah k paling optimal adalah k = 6. Berikut pembagian tiap item pada masing-masing pengelompokkan k=6 dapat dilihat pada gambar 3.



Gambar 3. Perbandingan Item Tiap *Cluster*

Berikut detail hasil pengelompokan daerah rawan bencana tanah longsor menggnakan algoritma *K-Means* dengan jumlah k paling optimal adalah k = 6 dapat dilihat pada tabel 2.

Tabel 2. Detail Hasil Pengelompokan Menggunakan Algoritma *K-Means* dengan k = 6

Kelompok/ Kluster	Jumlah Daerah	Anggota Kelompok
0	4	Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Tasikmalaya, Kota Tasikmalaya
1	2	Kabupaten Sukabumi, Kota Bogor
2	12	Kabupaten Cianjur, Kabupaten Cirebon, kabupaten Karawang, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Subang, Kota bandung, Kota Banjar, Kota Bekasi, Kota Cimahi, Kota Depok, Kota Sukabumi
3	1	Kabupaten Bogor
4	1	Kabupaten Sumedang
5	4	Kabupaten Ciamis, Kabupaten Garut, Kabupaten Kuningan, Kabupaten Majalengka

Dapat dilihat pada tabel diatas bahwa kluster 2 merupakan kluster dengan jumlah daerah paling banyak. Sedangkan, pada gambar 3 diatas dapat diketahui bahwa jumlah kejadian terbanyak terletak pada kluster 5 dengan jumlah daerah adalah 4 daerah dan jumlah kejadian sebanyak 106 kejadian.

4. KESIMPULAN

Dapat disimpulkan bahwasannya pengelompokkan dengan menggunakan metode *K-Means* lebih optimal dibandingkan dengan mengguakan metode *K-Medoids* pada data kejadian tanah longsor Provinsi Jawa Barat pada tahun 2019 dengan jumlah k paling optimal adalah k = 6. Perolehan *cluster* dominan, menunjukkan bahwa kluster 2 merupakan kluster dengan jumlah daerah paling banyak. Dan jumlah kejadian terbanyak terletak pada kluster 5 dengan jumlah daerah adalah 4 daerah dan jumlah kejadian sebanyak 106 kejadian.

REFERENSI

[1] D. Setianingsih and R. F. Hakim, "Penerapan *Data Mining* dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan Association Rule Algoritma Apriori," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015*, pp. 731–741, 2015.

[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] 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.

[4] T. Velmurugan, "Efficiency of *k-Means* and *K-Medoids* Algorithms for *Clustering* Arbitrary Data Points," *Int. J. Comput. ...*, vol. 3, no. 5, pp. 1758–1764, 2012, [Online]. Available: http://www.researchgate.net/publication/233986697_Efficiency_of_*k-Means*_and_*K-Medoids*_Algorithms_for_*Clustering*_Arbitrary_Data_Points/file/d912f50dc62a03083a.pdf.

[5] S. Gultom, S. Sriadhi, M. Martiano, and J. Simarmata, "Comparison analysis of *K-Means* and K-Medoid with Ecluidience Distance Algorithm, Chanberra Distance, and Chebyshev Distance for Big Data *Clustering*," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012092.

[6] S. Shah and M. Singh, "Comparison of a time efficient modified k-mean algorithm with K-mean and K-Medoid algorithm," *Proc. - Int. Conf. Commun. Syst. Netw. Technol. CSNT 2012*, pp. 435–437, 2012, doi: 10.1109/CSNT.2012.100.

[7] M. Tiwari and R. Singh, "Comparative Investigation of *K-Means* and K-Medoid Algorithm on Iris Data," *Int. J. Eng. Res. Dev.*, vol. 4, no. 8, pp. 69–72, 2012, [Online]. Available: www.ijerd.com.

[8] Noviyanto, "Penerapan *Data Mining* Dalam Mengelompokkan Jumlah Kematian," *Paradig. Inform. dan Komput.*, vol. 22, no. 2, 2020.

[9] M. Gading Sadewo, A. Perdana Windarto, and A. Wanto, "KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) PENERAPAN ALGORITMA *CLUSTERING* DALAM MENGELOMPOKKAN BANYAKNYA DESA/KELURAHAN MENURUT UPAYA ANTISIPASI/ MITIGASI BENCANA ALAM MENURUT PROVINSI DENGAN *K-MEANS*," vol. 2, pp. 311–319, 2018, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/komik.

[10] T. Velmurugan and T. Santhanam, "Computational complexity between *K-means* and *K-Medoids clustering* algorithms for normal and uniform distributions of data points," *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368, 2010, doi: 10.3844/jcssp.2010.363.368.

[11] S. Kasusdi, P. T. Rumah, and S. Padjadjaran, "Information System Journal," pp. 1–10, 2010.

[12] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "*K-Means* and *K-Medoids*: *Cluster* Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.

[13] P. Arora, Deepali, and S. Varshney, "Analysis of *K-Means* and *K-Medoids* Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.

[14] N. Arbin, N. S. Suhaimi, N. Z. Mokhtar, and Z. Othman, "Comparative analysis between *k-means* and *K-Medoids* for statistical *clustering*," *Proc. - AIMS 2015, 3rd Int. Conf. Artif. Intell. Model. Simul.*, pp. 117–121, 2016, doi: 10.1109/AIMS.2015.82.

[15] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid *clustering* x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.