ISSN 2460-1160

KLASTERISASI WILAYAH TANAH LONGSOR BERDASARKAN DAMPAK WILAYAH DAN GEOGRAFIS MENGGUNAKAN METODE K-MEANS

(Studi Kasus : Kabupaten dan Kota di Jawa Timur)

Moch. Rizki Eko Waluyo[1], Pramana Yoga Saputra.[2], Habibie Ed Dien.[3]

[1,2,3] Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
[1]mochrizkiekow@gmail.com, [2]pramana.yoga@polinema.ac.id, [3]habibie@polinema.ac.id

***Abstrak*— Tanah Longsor adalah suatu peristiwa geologi yang terjadi karena pergerakan massa batuan atau tanah dengan berbagai tipe dan jenis seperti jatuhnya bebatuan atau gumpalan besar tanah. Berdasarkan data yang dihimpun BPBD sepanjang tahun 2017-2018, terjadi sebanyak 212 kejadian dan mengakibatkan lebih dari 60 korban jiwa pada wilayah kabupaten & kota di Jawa Timur. Oleh karena itu, diperlukan usaha mengurangi dampak yang ditimbulkan dengan melakukan klasterisasi terhadap daerah kerawanan di wilayah kabupaten & kota di Jawa Timur. Dalam penelitian ini memanfaatkan teknik data mining dengan menggunakan metode *K-Means* untuk mengelompokkan wilayah terdampak tanah longsor dan geografis dan menggunakan metode *Pearson Correlation Coefficient* untuk mengetahui tingkat keterkaitan korelasi antara dampak wilayah tanah longsor dengan geografis. Berdasarkan hasil klasterisasi menggunakan metode *K-means* terdapat tiga klaster untuk data wilayah tanah longsor dan data geografis. Hasil pengujian mendapatkan nilai baik/optimal sebesar 0.021298251540091 untuk pengujian klasterisasi data wilayah tanah longsor dan nilai sebesar 0,326936115 untuk pengujian klasterisasi data geografis menggunakan *davies-bouldin index*. Sedangkan untuk hasil pengujian menggunakan *purity* hasil klasterisasi data wilayah tanah longsor mendapatkan nilai baik sebesar 0.92 dengan *error cluster* 0.08 dan hasil klasterisasi data geografis mendapatkan nilai baik sebesar 0.92 dengan *error cluster* 0.08 sehingga kedua klasterisasi dapat dikatakan bagus/optimal. Hasil pengukuran keterkaitan data menggunakan *Pearson Correlation Coefficient* mendapatkan nilai 2.1700691137793E-7 nilai korelasi positif yang berarti data dampak wilayah tanah longsor dan geografi memiliki korelasi secara linear yang berarti semakin tinggi nilai data geografi maka semakin tinggi pula nilai data dampak wilayah tanah longsor. Kesimpulan dari penelitian ini adalah metode *K-means* dapat digunakan untuk melakukan pengelompokkan wilayah terdampak tanah longsor dan geografis.**

Kata kunci—data mining, klasterisasi, longsor, korelasi

I. PENDAHULUAN

Bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan, baik oleh faktor alam dan atau faktor non alam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis [1]. Bencana merupakan pertemuan dari tiga unsur, yaitu ancaman bencana, kerentanan, dan kemampuan yang dipicu oleh suatu kejadian [2], dengan definisikan bencana menjadi beberapa jenis tentunya memerlukan penanganan yang berbeda pula.

Bencana tanah longsor adalah bencana yang sering terjadi di berbagai wilayah di indonesia dan dapat mengakibatkan korban jiwa karena kejadian dan waktu tiba-tiba. Jawa timur merupakan salah satu dari tiga provinsi yang sering terjadi bencana tanah longsor di indonesia. Bencana tanah longsor hampir merata di setiap daerah kabupaten dan kota di provinsi jawa timur. Berdasarkan data yang dihimpun oleh Badan Nasional Penanggulangan Bencana (BNPB), Pada Provinsi Jawa Timur kejadian tanah longsor sepanjang tahun 2017-2018 terjadi sebanyak 212 kejadian, dari total kejadian tersebut kabupaten ponorogo dan trenggalek menjadi daerah dengan jumlah kejadian tertinggi di Provinsi Jawa Timur. Kejadian tersebut juga mengakibatkan lebih dari 60 korban jiwa, sekitar 1998 rumah rusak berat dan sekitar 20 fasilitas pendidikan rusak.

Badan Penanggulangan Bencana Daerah (BPBD) merupakan salah satu lembaga pemerintah yang menjadi pelaksana penanggulangan bencana daerah di Indonesia khususnya di Kabupaten/Kota dan Provinsi. Dalam melakukan penanganan bencana BPBD menghasilkan data, data tersebut berupa data pascabencana/data wilayah terdampak tanah longsor. Peneliti berencana memanfaatkan data tersebut untuk dilakukan analisa terhadap metode yang digunakan dalam penelitian ini. Hal ini dilakukan agar BPBD dalam hal ini sebagai pelaksana penanggulangan

bencana di daerah dapat melakukan langkah awal (Pra-Bencana) guna mengurangi dampak kerusakan dari kejadian tanah longsor sebelumnya.

Pada Penelitian ini menggunakan analisis metode korelasional dan klasterisasi menggunakan data dari Badan Penanggulangan Bencana Daerah (BPBD) Jawa Timur Tahun 2017-2019 dan beberapa literatur sumber terkait penelitian ini. Dari data tersebut diperoleh data dampak wilayah tanah longsor dan geografis. Dimana parameter data dampak wilayah adalah jumlah kejadian, kerusakan dan korban jika sedangkan data geografis berupa kemiringan lereng tanah,jenis tanah,dan curah hujan tahunan. Selanjutnya, hasil dari klaster tersebut dianalisis menggunakan menggunakan metode korelasional untuk menemukan korelasi dari data tersebut.

Peneliti berharap dengan adanya penelitian ini diharapkan dapat meningkatkan kinerjapihak BPBD untuk mengolah data-data wilayah tanah longsor dan menganalisis keterkaitan dampak wilayah dan geografis di Provinsi Jawa Timur. Oleh karena itu,berkaitan dengan adanya penelitian ini peneliti berencana membuat sebuah sistem pengelompokkan data wilayah tanah longsor dan menganalisis keterkaitan dampak wilayah tanah longsor dan geografis di Provinsi Jawa Timur.

II. LANDASAN TEORI

A. *Data Mining*

Data mining adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah besar data. Sumber data dapat mencakup basis data, data gudang, Web, repositori informasi lain, atau data yang dialirkan ke Internet secara dinamis [3]. Data mining juga disebut sebagai proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data mining.

B. *Clustering*

Clustering atau Klasterisasi adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek di kelompok lain. Clustering sendiri juga disebut sebagai Unsupervised Classification, karena clustering lebih bersifat untuk dipelajari dan diperhatikan. Clustering berbeda dengan klasifikasi, dalam hal tidak ada variabel target. Algoritma-algoritma clustering digunakan untuk menentukan segmen keseluruhan himpunan data menjadi sub grup yang relatif sama atau cluster dengan kesamaan record dalam cluster dimaksimumkan dan kesamaan record di luar cluster diminimumkan [4]. Oleh karena itu, Clustering sangat berguna dan bisa menemukan grup yang tidak dikenal dalam data.

C. *K-means*

Algoritma *K-Means* adalah salah satu algoritma dalam clustering untuk mengelompokkan data dengan atribut tertentu ke dalam bentuk satu atau lebih kelompok yang diinisialisasi sebanyak k. Dalam algoritma *k-means* langkah awal yang dilakukan adalah menentukan k objek secara acak sebagai pusat data (Centroid), kemudian dihitung jarak antara setiap data dengan setiap pusat data (Centroid). Selanjutnya, penempatan data ke dalam pusat data (Centroid) yang terdekat dan melakukan iterasi sampai cluster tidak berubah (konvergen). Berikut merupakan tahapan-tahapan yang dilakukan dalam mengelompokan data dengan menggunakan metode *k-means* [5]:

1. Tentukan jumlah klaster (k),tetapkan pusat klaster sembarang.
2. Hitung jarak setiap data ke pusat klaster menggunakan *Euclidean Distance Formula* ditunjukkan ke dalam persamaan berikut:

$$d(x,y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \quad (1)$$

Keterangan :

d(x,y) = jarak data ke x ke pusat klaster y.
x_i = data x pada observasi ke-i.
y_i = titik pusat ke y observasi ke-i.
n = banyaknya observasi,

3. Mengelompokkan data yang menjadi anggota pada setiap klaster berdasarkan kedekatannya dengan centroid.
4. Hitung pusat klaster yang baru dengan men.cari rata-rata dari data-data yang menjadi anggota pada klaster tersebut.

$$C_{ij} = \frac{\sum_{i=1}^{p} x_{ij}}{p} \quad (2)$$

Keterangan :

Cij = centroid terbaru pada iterasi k.
X_{ij} = anggota cluster ke k.
P = banyaknya anggota cluster ke k

Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke klaster yang lain.

D. *Davies-Bouldin Index*

Dalam pengujian hasil klasterisasi yang pertama menggunakan metode *Davies-Bouldin Index* (DBI). DBI merupakan metode yang digunakan untuk menguji validitas jumlah klaster. DBI ini memaksimalkan jarak inter-cluster antara cluster Ci dan Cj dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah cluster [6]. Berikut merupakan tahapan-tahapan dalam perhitungan Davies-Bouldin Index :

1. Data yang digunakan merupakan data hasil klasterisasi dari proses k-means.
2. Cari nilai *Sum of Square Within* (SSW) untuk setiap cluster.
3. Cari nilai *Sum of Square Between-Cluster* (SSB).
4. Cari nilai *Ratio* (R).

$$R_{(i,j_{..}n)} = \frac{SSW_i + SSW_j + .. SSW_n}{SSB_{i,j} + .. + SSB_{n,n}} \quad (3)$$

Semakin rendah nilai r (mendekati 0 dan tidak negatif), semakin baik kualitas cluster [7].

E. *Purity*

Dalam pengujian hasil klasterisasi yang kedua, menggunakan metode purity. *Purity* salah satu metode uji cluster dengan semua objek class yang sama berada pada cluster yang sama dikatakan murni (pure) [8]. Berikut rumus untuk menghitung nilai purity :

$$r = \frac{1}{n}\sum_{i=1}^{k} a_i \quad (4)$$

Keterangan :

r = tingkat akurasi cluster purity.
k = jumlah cluster.
a = objek yang muncul.

Semakin tinggi nilai r (semakin mendekati 1), semakin baik kualitas cluster [9]. Untuk menghitung error cluster pada purity dapat menggunakan persamaan berikut:

$$e = 1 - r \quad (5)$$

Keterangan :

e = tingkat error cluster.
r = tingkat akurasi cluster purity.

F. *Pearson-Correlations Coefficient*

Pearson's Correlation Coeffiecient adalah salah satu metode uji statistik multikolinearitas yang mengukur keterkaitan antara dua atau lebih variabel [10], nilai *Pearson's Correlation Coeffiecient* ini lah yang akan melandasi untuk menentukan pengukuran seberapa kuat hubungan antarvariabel.. Berikut perhitungan yang digunakan untuk menghitung nilai pearson :

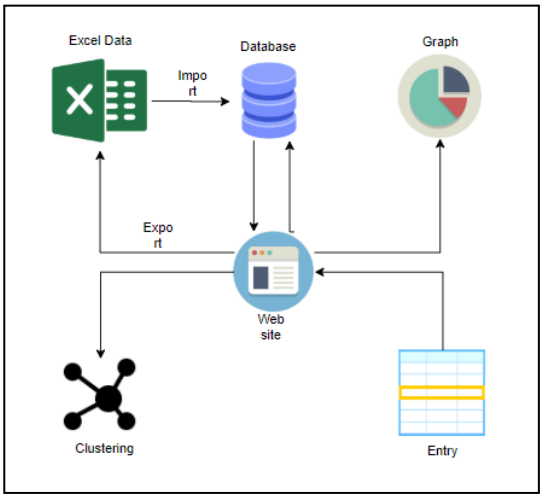
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (6)$$

Keterangan :

r = koefiesien korelasi pearson.
n = jumlah data.
x = variabel ke-x.
y = variabel ke-y.

III. ANALISA DAN PERANCANGAN

A. *Analisis Sistem*



Gambar 1. Gambaran Umum Sistem

Sistem akan memetakan daerah terdampak tanah longsor dan geografis berdasarkan hasil klasterisasi dari data dokumen kejadian bencana tanah longsor pada tahun 2017-2019 dan data geografis. Masukan sistem berupa data terdampak tanah longsor dan geografis di kabupaten/kota yang tercatat pihak Badan Penanggulangan Bencana Daerah. Masukan tersebut akan diolah menggunakan teknik klasterisasi dengan metode *k-means* dan akan menghasilkan keluaran berupa tabel dan grafik yang sesuai dengan kluster yang terbentuk. Setelah dilakukan proses *k-means*, lalu dilanjutkan dengan melakukan korelasi pada data. Output yang dihasilkan dari korelasi yang akan melandasi untuk menentukan pengukuran hubungan antarvariabel/antardata.

B. *Analisis Kebutuhan*

Analisis kebutuhan merupakan tahap untuk menganalisis segala sesuatu yang dibutuhkan untuk mengembangkan sistem, dalam hal ini terdapat beberapa hal yang harus dipenuhi antara lain meliputi dataset.

Data yang digunakan adalah data dokumen kejadian bencana tanah longsor dan geografis di kabupaten/kota yang tercatat di Badan Penanggulangan Bencana Daerah (BPBD). Adapun atribut dari data kejadian tanah longsor dan geografis yang akan dilibatkan dalam proses klasterisasi adalah sebagai berikut :

Tabel I. Keterangan Data Wilayah Terdampak Tanah Longsor

No	Atribut Parameter	Keterangan
1	Jumlah Kejadian	Jumlah Kejadian tanah longsor
2	Jumlah Korban Jiwa	Akumulasi jumlah korban jiwa dari korban meninggal, luka-luka, dan terdampak
3	Jumlah Kerusakan	Akumulasi jumlah kerusakan dari kerusakan rumah rusak

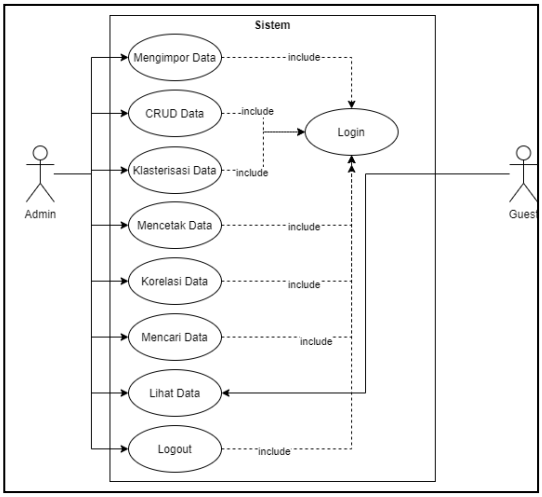
Tabel II. Keterangan Data Geografis

No	Atribut Parameter	Keterangan
1	Curah Hujan	Rerata Curah Hujan Tahunan di daerah Kabupaten dan Kota
2	Kemiringan Lereng	Rerata Kemiringan Tanah di daerah Kabupaten dan Kota berdasarkan ketinggian topografi dibagi jarak datar sebenarnya
3	Jenis Tanah	Rerata Jenis Tanah di daerah Kabupaten dan Kota
4	Penggunaan Lahan Tegal	Total penggunaan lahan tegal di daerah Kabupaten dan Kota
5	Penggunaan Lahan Huma	Total penggunaan lahan huma di daerah Kabupaten dan Kota
6	Penggunaan Lahan Sementarara tidak digunakan	Total lahan bukan sawah yang sementara tidak diusahakan di daerah Kabupaten dan Kota

C. *Use Case Diagram*

Pengguna dalam sistem ini ada dua yaitu guest dan admin. Interaksi yang dapat dilakukan oleh admin dengan

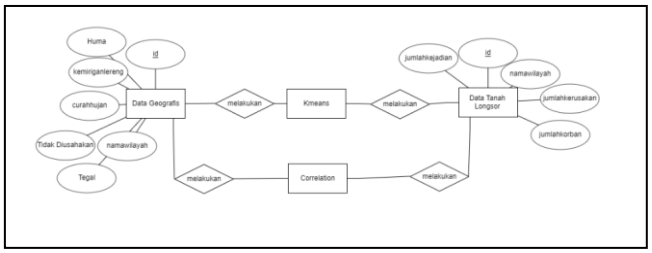
sistem ini adalah mengelola data dampak wilayah tanah longsor dan geografis, melakukan proses klasterisasi, melihat hasil klasterisasi, melihat pemetaan hasil klasterisasi dan melakukan korelasi. sedangkan user hanya dapat melihat data. Berikut merupakan use case diagram pada sistem klasterisasi terlihat pada Gambar 2 :



Gambar 2. Use Case Diagram

D. *Entity Relationship Diagram*

Entity relationship diagram merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. Berikut *entity relationship diagram* untuk sistem yang akan dibangun pada Gambar 3:

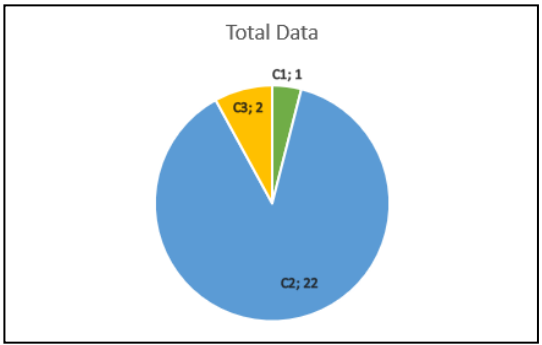


Gambar 3. Entity Relationship Diagram

IV. HASIL DAN PENGUJIAN

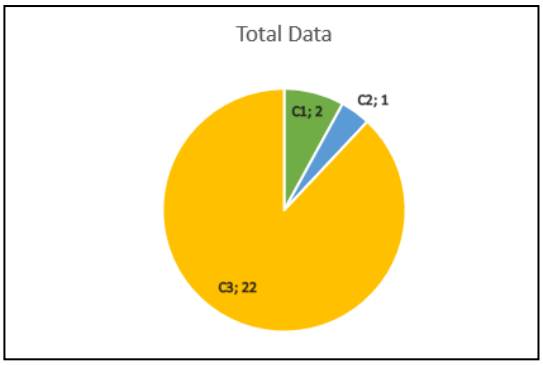
A. *Hasil*

Dari hasil analisa menggunakan metode *k-means* yang ditunjukkan gambar 4 pada data wilayah terdampak tanah longsor didapati untuk cluster 1 terdapat 1 anggota cluster yang didominasi wilayah yang sering terdampak dan kerusakan parah,untuk cluster 2 terdapat 22 anggota cluster yang didominasi dengan wilayah yang mengalami kerusakan dan korban jiwa sedang dan untuk cluster 3 terdapat 2 anggota cluster didominasi wilayah yang mengalami kerusakan, kejadian dan korban jiwa rendah.



Gambar 4. Pie Chart Hasil Klasterisasi Data Wilayah Terdampak

Hasil analisa menggunakan metode K-means ditunjukkan gambar 5 pada data geo didapati untuk cluster 1 terdapat 2 anggota cluster,untuk cluster 2 terdapat 1 anggota cluster dan untuk cluster 3 terdapat 22 anggota cluster.



Gambar 5. Pie Chart Hasil Klasterisasi Data Geografis

B. *Pengujian*

Untuk mengetahui hasil dari metode yang telah diterapkan pada sistem maka dilakukan proses pengujian dengan keseluruhan data yaitu sejumlah 25 data wilayah terdampak tanah longsor dan geografis di kabupaten dan kota di Jawa Timur. Dalam pengujian klasterisasi yang dilakukan peneliti menggunakan 2 metode yaitu *Davies-Bouldin Index* dan *Purity* sedangkan untuk menguji korelasi kedua data yang digunakan menggunakan *Pearson Correlation Coefficient*.

1) *Pengujian Metode K-Means Dacies-Bouldin Index*

Uji coba ini dilakukan untuk menguji seberapa baik pengelompokan telah dilakukan dibuat menggunakan jumlah dan fitur yang melekat pada dataset setelah dilakukan proses clustering dengan menggunakan metode *k-means clustering*. Dari hasil proses *k-means* akan terlihat hasil iterasi terakhir, selanjutnya dari iterasi tersebut akan dihitung *sum of square within* dan *sum of square between cluster*, berikutnya akan dihitung ratio menggunakan rumus (1). Dari hasil pengujian tersebut pada klasterisasi menunjukkan bahwa sistem dapat melakukan klasterisasi data secara baik karena nilai DBI yang diperoleh mendekati nilai nol, untuk hasil pengujian terlihat pada Tabel III:

Tabel III. Hasil Pengujian Davies-Bouldin Index

Data	**Hasil Pengujian Davies-Bouldin Index**	
	Hasil	***Keterangan***
Wilayah Terdampak	0,021298251540091	Baik
Geografis	0,326936115	Baik

2) *Pengujian Metode K-Means Purity*

Uji coba ini dilakukan untuk menguji seberapa baik pengelompokan telah dilakukan dinilai berdasarkan data yang tidak digunakan untuk pengelompokan, seperti label kelas yang dikenal dan tolok ukur eksternal pada dataset setelah dilakukan proses clustering dengan menggunakan

metode *k-means clustering*. Dari hasil proses k-means akan terlihat hasil iterasi terakhir, selanjutnya dari iterasi tersebut akan dihitung obyek sama pada cluster, berikutnya akan dihitung ratio menggunakan rumus (2). Dari hasil pengujian pada hasil klasterisasi menunjukkan bahwa sistem dapat melakukan klasterisasi data secara cukup baik karena purity dikatakan optimal/baik jika bernilai 1 atau mendekati 1, untuk hasil pengujian terlihat pada Tabel IV :

Tabel IV. Hasil Pengujian Purity

Data	**Hasil Pengujian Purity**		
	Hasil	***Error Cluster***	***Keterangan***
Wilayah Terdampak	0,92	0,08	Baik
Geografis	0,92	0,08	Baik

3) Pengujian Korelasi Pearson's Correlation Coefficient

Uji coba ini dilakukan untuk menguji korelasi data yang digunakan. Dari hasil pengujian akan terlihat hasil nilai korelasi yang menentukan seberapa kuat hubungan antarvariabel/antardata. Dari hasil pengujian korelasi menunjukkan bahwa sistem dapat melakukan korelasi data secara cukup baik dengan menghasilkan nilai korelasi positif dan bermakna semakin tinggi nilai data geografi maka semakin tinggi pula nilai data dampak wilayah tanah longsor, untuk hasil pengujian terlihat pada Tabel V:

Tabel V. Hasil Pengujian Korelasi

Data	**Hasil Pengujian Korelasi**	
	Hasil	***Keterangan***
Wilayah Terdampak dan Geografis	2.1700691137793E-7	Positif

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil pengujian dan analisis penelitian terhadap pemetaan daerah wilayah terdampak tanah longsor dengan menggunakan metode k-means, maka dapat diambil beberapa kesimpulan yaitu :

1. Metode *k-means* dapat digunakan untuk melakukan pengelompokan data wilayah terdampak dan geografis di kab/kota Jawa Timur.
2. Berdasarkan hasil pengujian mendapatkan nilai baik/optimal sebesar 0.021298251540091 untuk pengujian hasil klasterisasi data wilayah tanah longsor dan nilai sebesar 0,326936115 untuk pengujian hasil klasterisasi data geografis menggunakan *davies-bouldin index*.
3. Berdasarkan hasil pengujian menggunakan *purity* hasil klasterisasi data wilayah tanah longsor mendapatkan nilai baik sebesar 0.92 dengan error cluster 0.02, Sedangkan hasil klasterisasi mendapatkan nilai baik sebesar 0.92 dengan error cluster 0.08.
4. Berdasarkan perhitungan korelasi yang telah dilakukan terhadap data dampak wilayah tanah longsor dan data geografis dengan metode *Pearson Correlation Coefficient* diperoleh nilai korelasi PCC sebesar 2.1700691137793E-7 dengan nilai korelasi positif berarti data dampak wilayah tanah longsor dan geografi memiliki korelasi secara linear.

B. Saran

Dalam pengerjaan penelitian ini terdapat beberapa kekurangan yang dapat diperbaiki dan ditingkatkan lagi untuk penelitian selanjutnya. Oleh karena itu, berikut beberapa saran dan masukan yang diharapkan dapat memberikan perbaikan dalam penelitian selanjutnya, antara lain :

1. Untuk meningkatkan hasil akurasi *clustering*, dapat ditambahkan metode penentuan centroid awal dan penentuan jumlah klaster awal agar menghasilkan clustering yang lebih akurat.
2. Hasil klasterisasi dapat divisualisasikan dalam bentuk *geomap* dengan ditambahkan atribut *longitude* dan *latitude*.

DAFTAR PUSTAKA

[1] BNPB., Definisi Bencana: bnpb. Dipetik 2019, dari bnpb: https://bnpb.go.id/definisi-bencana, 2019.

[2] Indonesia., Undang-Undang Nomor 24 Tahun 2007 : Tentang Penanggulangan Bencana. Jakarta, 2007.

[3] Han, J. M., & Pei., Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.

[4] Larose, D. T., Discovering Knowledge in Data: An Introduction to Data Mining, 2005, Wiley-Interscience.

[5] Shreya, B., Ankit, C., & Somnath, P., "Empirical Evaluation of K-Means, Bisecting KMeans,Fuzzy C-Means and Genetic K-Means.", 2015, https://ieeexplore.ieee.org/document/7443889.

[6] Wani, M. A., & Riyaz, R., "A novel point density based validity index for clustering gene.", 2017, doi:10.1504/IJDMB.2017.084027.

[7] Santoso, H., & Magdalena, H., "Improved K-Means Algorithm on Home Industry Data Clustering in the Province of Bangka Belitung", 2020, doi:10.1109/ICoSTA48221.2020.1570598913.

[8] Rohmawati, N., Sofi, D., & Mohamad, J., "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa", 2015.

[9] Stanford., Diambil kembali dari nlp.standford.edu: https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html, 2009.

[10] Haomiao, Z., Zhihong, D., Yuanqing, X., & Mengyin, F., "A new sampling method in particle filter based on Pearson correlation coefficient", 2016.