

PENGELOMPOKAN DATA DENGAN MENGGUNAKAN METODE FUZZY C-MEANS UNTUK WILAYAH RAWAN BENCANA TANAH LONGSOR

Faisal Akbar

Sekolah Tinggi Ilmu Komputer (STIKOM) Poltek Cirebon Jawa Barat, Indonesia

Email: faisal.akbar@stikompoltek.ac.id

INFO ARTIKEL	ABSTRAK
Diterima 02 November 2020 Diterima dalam bentuk revisi 17 November 2020 Diterima dalam bentuk revisi	Bencana tanah longsor merupakan salah satu bencana alam yang sering terjadi di beberapa daerah di Indonesia. Kabupaten Kuningan salah satu daerah dengan lokasi dataran yang secara geografis merupakan daerah yang memiliki permukaan tanah relatif datar dengan variasi berbukit-bukit menyebabkan beberapa wilayah di Kabupaten Kuningan sangat rawan terjadinya longsor saat musim hujan. Dalam penanganan pra-bencana di wilayah tersebut, pemerintah dan BPBD (Badan Penanggulangan Bencana Daerah) Kabupaten Kuningan melakukan mitigasi bencana yang bertujuan untuk meminimalisir kemungkinan terjadinya longsor. Perlu adanya sistem yang lebih terstruktur sebagai upaya mitigasi bencana longsor, dengan mengolah data kejadian bencana menjadi suatu sistem informasi penanganan pra-bencana. Pengolahan data bencana alam dalam penelitian ini yaitu menggunakan teknik <i>data mining</i> , data yang diolah dengan teknik ini kemudian menghasilkan suatu sistem informasi baru yang bersumber dari data lama. <i>Clustering</i> atau pengelompokan data merupakan salah satu konsep penganalisaan <i>Data Mining</i> dengan metode penganalisaan proses partisi satu <i>set</i> objek data ke dalam himpunan bagian yang disebut dengan <i>cluster</i> . Algoritma yang digunakan adalah <i>Fuzzy C-Means</i> (FCM). Algoritma ini memberi kebebasan dalam hal jumlah <i>cluster</i> yang akan dibuat, juga memiliki tingkat akurasi yang tinggi.
Kata kunci: Pengelompokan Data; Metode <i>Fuzzy C-Means</i> ; Bencana Longsor	

Pendahuluan

Tanah longsor biasa terjadi pada musim hujan di dataran tinggi atau pegunungan, tetapi longsor juga dapat terjadi pada dataran yang relatif rendah. Kabupaten Kuningan dengan lokasi dataran yang secara geografis merupakan daerah dengan dataran yang memiliki permukaan tanah relatif datar dengan variasi berbukit-bukit menyebabkan beberapa wilayah di Kabupaten Kuningan sangat rawan terjadinya longsor saat musim hujan.

Informasi tentang jumlah kejadian bencana longsor sangat dibutuhkan untuk penanggulangan bencana di Kabupaten Kuningan. Perlu adanya sistem yang lebih terstruktur sebagai upaya mitigasi bencana longsor, yaitu dengan mengolah data tersebut menjadi suatu sistem informasi penanganan pra-bencana. Pengolahan data bencana alam yang dilakukan peneliti yaitu menggunakan teknik *Data Mining*, karena metode ini dapat menjadi langkah awal dalam kegiatan mitigasi bencana melalui teknik penanganan non- fisik dengan menganalisa data historis dari data bencana longsor yang ada. Dari banyaknya konsep penganalisaan *Data Mining*, *Clustering* atau klasterisasi atau pengelompokan yakni salah satu metode penganalisaan dengan proses partisi satu *set* objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Partisi pada proses *clustering* tidak dilakukan secara manual melainkan dengan suatu algoritma (Pamusti, 2018). Algoritma yang akan digunakan untuk penelitian ini adalah *Fuzzy C-Means* (FCM) (Bifawaidati, 2017). Algoritma *Fuzzy C-Means* memberi kebebasan dalam hal jumlah *cluster* yang akan dibuat. Kelebihan lainnya adalah ia dapat melakukan *clustering* lebih dari satu variabel secara sekaligus. Data-data beserta parameternya dapat dikelompokkan dalam beberapa *cluster* sesuai dengan kecenderungannya. *Fuzzy C-Means* juga memiliki tingkat akurasi yang tinggi dan waktu komputasi yang cepat (Arista et al., 2017). diharapkan sistem informasi terkait dapat dijadikan acuan untuk antisipasi dan kewaspadaan dalam proses mitigasi peristiwa longsor serta dapat menjadi solusi atas permasalahan penanggulangan bencana tanah longsor di wilayah tersebut.

Metode Penelitian

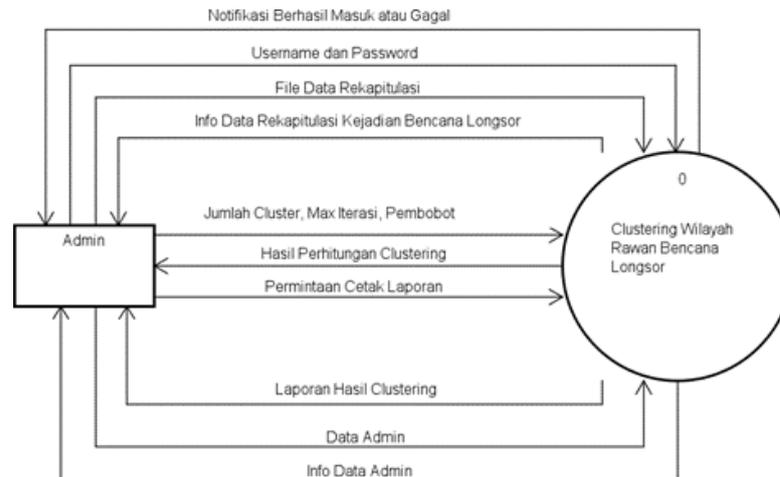
Metode yang digunakan peneliti adalah metode kuantitatif. Karena penelitian ini menggunakan sample data berupa angka-angka dan penekanan pada pengukuran hasil yang objektif menggunakan analisis statistik. Fokus metode kuantitatif adalah mengumpulkan data dan melakukan generalisasi untuk menjelaskan fenomena khusus yang dialami oleh populasi. Tujuannya membentuk hipotesis dan mendeskripsikan suatu fenomena. Jika suatu hipotesis lolos uji berkali-kali, hipotesis tersebut dapat menjadi suatu teori ilmiah baru.

Metode kuantitatif dapat diartikan sebagai metode penelitian yang berlandaskan pada filsafat positivisme, digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik, dengan tujuan untuk menguji hipotesis yang telah ditetapkan (Sugiyono, 2016).

Hasil dan Pembahasan

A. Hasil Penelitian

Pada penelitian ini menghasilkan sebuah sistem dengan alir data yang digambarkan pada digram konteks dan dapat dilihat pada Gambar 1 sesuai dengan permasalahan yang telah dirumuskan.



Gambar 1 Diagram Konteks

Berdasarkan dari diagram konteks pada Gambar 1, maka dapat menghasilkan sebuah sistem yang memiliki desain prosedur sebagai berikut:

- 1) Prosedur *Input Data*
 - a) Admin login terlebih dahulu
 - b) Apabila login berhasil maka akan masuk ke dalam tampilan halaman utama. Jika proses login gagal maka sistem akan mengkonfirmasi bahwa login gagal.
 - c) Admin dapat melakukan olah data seperti menambah data admin, mengedit data admin dan menghapus data admin.
 - d) Admin mengupload data berupa file excel yang sudah di praproses data berisi rekap data kejadian bencana longsor 3 tahun terakhir per- kecamatan di Kabupaten Kuningan untuk dimasukkan kedalam data pengujian yang akan dikelompokkan.
- 2) Prosedur Pengelompokan Data Fuzzy C-Means
 - a) Admin menentukan jumlah klaster yang diinginkan.
 - b) Admin menentukan nilai maksimum iterasi.
 - c) Sistem akan membaca data yang dikelompokkan dari data pengujian dengan proses *clustering* menggunakan algoritma Fuzzy C-Means.
 - d) Sistem akan membaca jumlah kriteria yang sudah ditentukan.
 - e) Sistem membaca jumlah *cluster* yang diinput.
 - f) Sistem menentukan nilai bilangan random secara acak dengan nilai keanggotaan pada interval 0 sampai dengan 1.
 - g) Sistem menghitung fungsi objektif pada proses iterasi.
 - h) Jika proses melebihi maksimum iterasi maka proses Fuzzy C-Means berhenti.
 - i) Kemudian admin dapat menyimpan hasil *clustering* Fuzzy C-Means.
- 3) Prosedur Pencetakan Hasil Pengelompokan Data
 - a) Setelah data diproses, program akan menghasilkan nilai *Cluster* berdasarkan *Cluster* yang ditentukan, atribut kecamatan dan tahun kejadian.
 - b) Klik tombol 'Cetak Laporan' untuk mencetak hasil *clustering*.

B. Pembahasan

1) *Data Mining*

Data mining adalah analisis pengamatan *database* untuk menemukan hubungan yang tidak terduga dan untuk meringkas data dengan cara atau metode baru yang dapat dimengerti dan bermanfaat kepada pemilik data (Melilisa, 2011).

Konsep utama dari sebuah transformasi dalam *data mining* yakni suatu kumpulan data yang bersumber dari *database* yang berukuran besar yang diekstrak dan dirangkum untuk menemukan suatu pola berupa informasi yang bermanfaat dan mengandung sebuah pengetahuan yang dapat digunakan untuk membantu pengambil keputusan dalam mengambil kebijakan yang tepat untuk kepentingan sebuah organisasi.

2) Pengelompokan Data

Clustering atau pengelompokan data mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok (Syahputra et al., 2018). *Clustering* membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar diantara kelompok-kelompok. Metode *clustering* secara umum dapat dibagi menjadi dua, yaitu *hierarchical clustering* dan *partitional clustering* (Tan, 2017).

a) *Hierarchical Clustering*

Pada *hierarchical clustering* data dikelompokkan melalui suatu bagan yang berupa hirarki, dimana terdapat penggabungan dua grup yang terdekat disetiap iterasinya ataupun pembagian dari seluruh set data kedalam cluster. Contoh metode hierarchy clustering: Single Linkage, Complete Linkage, Average Linkage, Average Group Linkage.

b) *Partitional Clustering*

Partitional clustering yaitu data yang dikelompokkan ke dalam beberapa bagian menjadi sejumlah *cluster* tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada metode *partitional clustering* setiap *cluster* memiliki titik yang menjadi pusat *cluster* (Centroid) dan secara umum metode ini memiliki fungsi tujuan yaitu meminimumkan jarak (Dissimilarity) dari seluruh data ke pusat cluster masing-masing. Beberapa metode bagian dari *partitional clustering*: K-Means, Fuzzy C-means dan Mixture Modelling. Adapun manfaat dari *clustering* atau pengelompokan data, yaitu:

- 1) *Clustering* merupakan metode segmentasi data yang sangat berguna dalam analisis masalah bisnis tertentu. Misalnya segmentasi pasar, *marketing* dan pemetaan zona wilayah.
- 2) Identifikasi obyek dalam berbagai bidang seperti *computer vision* dan *image processing*.

3) Analisis Cluster

Analisis *Cluster Variance* digunakan untuk mengukur nilai hasil penyebaran data-data hasil *clustering* (Alfina et al., 2012). *Variance* pada *clustering* ada 2 (dua) jenis, yaitu:

- a) *Variance within cluster* merupakan jenis varian yang mengacu pada jarak antar anggota pada *cluster*.
- b) *Variance between cluster* merupakan jenis varian yang mengacu pada jarak antar *cluster*.

Untuk dapat melakukan analisis cluster ialah dengan cara melalui tahapan atau algoritma sebagai berikut:

- a) Mencari nilai varian sebuah *cluster* dengan cara dihitung dengan menggunakan rumus sebagai berikut:

$$V_c^2 = \frac{1}{nc - 1} \sum_{i=1}^{nc} (d_i - \bar{d}_i)^2$$

Keterangan:

V_c^2 = *variance* pada *cluster* ke c

$c = 1 \dots k$, dimana k = jumlah *cluster*

nc = jumlah data pada *cluster* c

d_i = data ke- i pada suatu *cluster*

\bar{d}_i = rata-rata dari data pada suatu *cluster*

- b) Setelah nilai *variance cluster* telah diperoleh, kemudian dapat mencari nilai *variance within cluster* dan *variance between cluster* dengan menggunakan rumus sebagai berikut:

- 1) *Variance within cluster*

$$V_w = \frac{1}{N - K} \sum_{i=1}^K (n_i - 1) \cdot V_i^2$$

Keterangan:

N = jumlah seluruh data

K = jumlah *cluster*

n_i = jumlah data pada *cluster* ke- i

V_i^2 = *variance* pada *cluster* ke- i

- 2) *Variance between cluster*

$$V_b = \frac{1}{N - K} \sum_{i=1}^K n_i (\bar{d}_i - \bar{d})^2$$

Keterangan:

\bar{d} = rata-rata dari \bar{d}_i

- c) Kemudian setelah memperoleh nilai variance within cluster dan variance between cluster, selanjutnya yaitu dengan menghitung nilai variance minimum dan variance maximum dengan rumus berikut:

$$V_{min} = \frac{V_w}{V_b}$$

$$V_{max} = \frac{V_b}{V_w}$$

4) Metode *Fuzzy C-Means*

Metode *Fuzzy C-Means* adalah teknik pembuatan *cluster* dimana keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaannya (Kemala et al., 2019). Berikut merupakan *Flowchart Fuzzy C-Means* yang dijelaskan pada Gambar 2.



Gambar 2 *Flowchart Fuzzy C-Means*

Algoritma Fuzzy C-Means adalah sebagai berikut:

- Input* data di *cluster* X, berupa matriks berukuran n x m (n=jumlah sample data, m=atribut setiap data). X_{ij} = data sample ke-i (i=1,2,...,n), atribut ke-j (j=1,2,...,m).
- Keterangan ketentuan nilai:
 - Jumlah cluster = c;
 - Pangkat = w;
 - Maksimum iterasi = MaxIter; Error terkecil (epsilon) = 0,00001;

Fungsi obyektif awal = $P_0 = 0$;

Iterasi awal = $t=1$;

- c. Bangkitkan nilai acak μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$; sebagai elemen-elemen matriks partisi awal u . μ_{ik} adalah derajat keanggotaan yang merujuk pada seberapa besar kemungkinan suatu data dapat menjadi anggota ke dalam suatu cluster. Posisi dan nilai matriks dibangun secara random. Dimana nilai keanggotaan terletak pada interval 0 sampai dengan 1. Hitung jumlah setiap kolom (atribut).

$$Q_j = \sum_{k=1}^c \mu_{ik}$$

Q adalah jumlah nilai derajat keanggotaan perkolom = 1 dengan $j=1,2,\dots,m$.

Hitung:

$$V_{kj} = \frac{\mu_{ik}}{Q_j}$$

- d. Hitung pusat Cluster ke- k : V_{kj} , dengan $k=1,2,\dots,n$; dan $j=1,2,\dots,n$.

e.
$$\mu_{ik} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

V = Nilai centroid atau pusat cluster didapat dari perhitungan total μ_{ik} kuadrat dikali X_{ij} (X atribut data) dibagi total μ_{ik} kuadrat.

- f. Hitung fungsi obyektif pada iterasi awal. Fungsi obyektif digunakan sebagai syarat perulangan untuk mendapatkan pusat cluster yang tepat. Sehingga diperoleh kecenderungan data untuk masuk ke cluster mana pada step akhir.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

- g. Tahap ini menghitung jumlah X_{ij} (data atribut) dikurangi nilai pusat cluster dikali nilai pembobot. Kemudian perbaiki derajat keanggotaan dengan menghitung nilai perubahan matriks partisi setiap data pada setiap cluster. Hitung perubahan matriks partisi: dengan: $i=1,2,\dots,n$; dan $k=1,2,\dots,m$.

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left([\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}} \right)}$$

Perhitungan perubahan matriks berfungsi untuk mengetahui hasil data yang masuk kedalam pengelompokan yang sudah ditentukan.

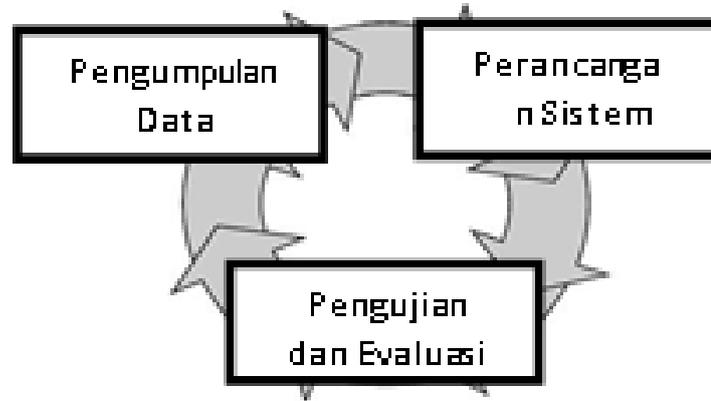
- h. Cek kondisi berhenti:

Jika: ($|P_t - P_{t-1}| < \epsilon$) atau ($t > \maxIter$) maka berhenti. Jika iterasi berhenti, ditentukan cluster dari tiap-tiap data. Cluster dipilih berdasarkan nilai matriks partisi terbesar.; jika tidak: $t=t+1$, ulangi langkah ke-4.

5) Metode *Prototype*

Model *Prototype* dimulai dari mengumpulkan spesifikasi kebutuhan hingga tahap implementasi, sehingga membantu tim pengembang perangkat

lunak untuk memahami lebih baik apa yang akan dikembangkan serta terbayang apa yang sebenarnya diinginkan. Adapun tahapannya dijelaskan melalui Gambar 3 sebagai berikut.



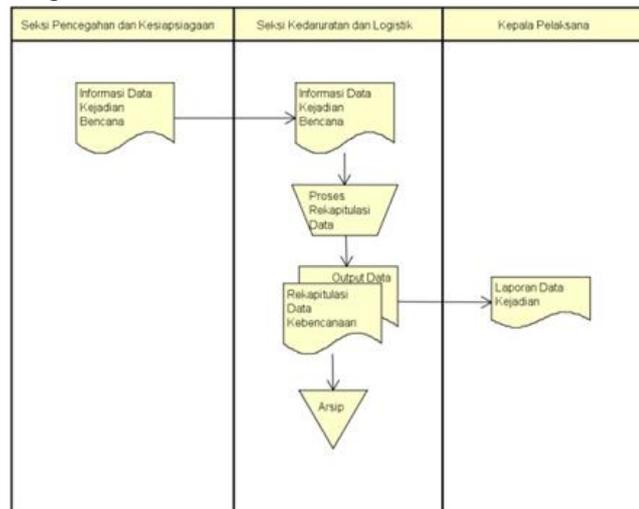
Gambar 3 Tahapan Model *Prototype*

Sumber: (Roger & Pressman, 2012)

- a. Pengumpulan Data
Pada tahap ini merupakan pengumpulan data yang terdiri dari data primer dan data sekunder. Pengumpulan data primer dihasilkan dari observasi studi kasus, kemudian melakukan wawancara dengan instansi. Sedangkan pengumpulan data sekunder dihasilkan dari dokumen observasi yang dilakukan dengan mempelajari dokumen-dokumen yang ada untuk memperoleh data dan informasi serta melakukan riset dengan merujuk pada buku-buku yang berhubungan dengan penelitian guna untuk memecahkan masalah pada penilaian ini.
 - b. Perancangan Sistem
Membangun sistem dengan membuat perancangan sementara yang berfokus kepada kebutuhan aplikasi dan alat bantu pemodelan.
 - c. Pengujian dan Evaluasi
Pada tahap ini dilakukan uji coba sistem yang telah dirancang dan mengevaluasi sistem untuk memastikan bahwa sistem sudah dapat digunakan dengan baik dan benar sesuai dengan yang diharapkan.
- 6) Analisis Sistem Berjalan
- Analisis prosedur sistem yang memberikan gambaran tentang sistem yang saat ini sedang berjalan. Analisis prosedur sistem bertujuan untuk mengetahui bagaimana cara kerja sistem tersebut sehingga kelebihan dan kekurangan sistem dapat diketahui.
- Berdasarkan pengamatan yang dilakukan di BPBD (Badan Penanggulangan Bencana Daerah) Kabupaten Kuningan mengenai Clustering Data Kejadian Bencana Longsor yaitu sistem yang berjalan pada BPBD Kabupaten Kuningan dalam pengelolaannya, informasi diperoleh berdasarkan data kejadian bencana yang dicatat oleh Seksi Pencegahan dan Kesiapsiagaan

yang selanjutnya catatan tersebut diserahkan kepada Seksi Kedaruratan dan Logistik sehingga data catatan tersebut diproses dan menjadi data rekapitulasi kejadian kebencanaan. Kemudian informasi data rekapitulasi tersebut diserahkan oleh Kepala Peaksana menjadi laporan data kejadian.

Berikut diagram sistem prosedur lama yang digunakan oleh BPBD Kabupaten Kuningan:



Gambar 4 Diagram *Flowmap* Prosedur Sistem Berjalan

7) Perancangan Sistem (Desain Prosedur)

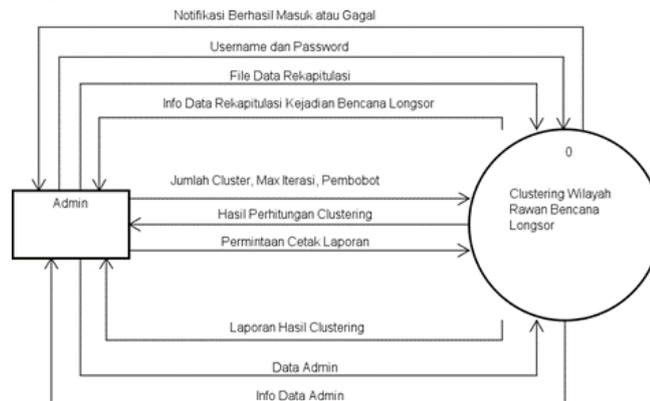
Terdapat beberapa prosedur sistem yang dibangun, yakni meliputi prosedur *input* data, prosedur *clustering fuzzy c-means*, dan prosedur mencetak hasil *clustering fuzzy c-means*, berikut penjelasan dari keseluruhan prosedur sistem tersebut:

- a. Admin login terlebih dahulu.
- b. Apabila login berhasil maka akan masuk ke dalam tampilan halaman utama. Jika proses login gagal maka sistem akan mengkonfirmasi bahwa login gagal.
- c. Admin dapat melakukan olah data seperti menambah data admin, mengedit data admin dan menghapus data admin.
- d. Admin mengupload data berupa file excel yang sudah di praproses data berisi rekap data kejadian bencana longsor 3 tahun terakhir per- kecamatan di Kabupaten Kuningan untuk dimasukkan kedalam data pengujian yang akan dikelompokkan.
- e. Admin menentukan jumlah klaster yang diinginkan.
- f. Admin menentukan nilai maksimum iterasi.
- g. Sistem akan membaca data yang dikelompokkan dari data pengujian dengan proses *clustering* menggunakan algoritma *fuzzy c-means*.
- h. Sistem akan membaca jumlah kriteria yang sudah ditentukan.
- i. Sistem membaca jumlah *cluster* yang diinput.

Pengelompokan Data Dengan Menggunakan Metode Fuzzy C-Means Untuk Wilayah Rawan Bencana Tanah Longsor

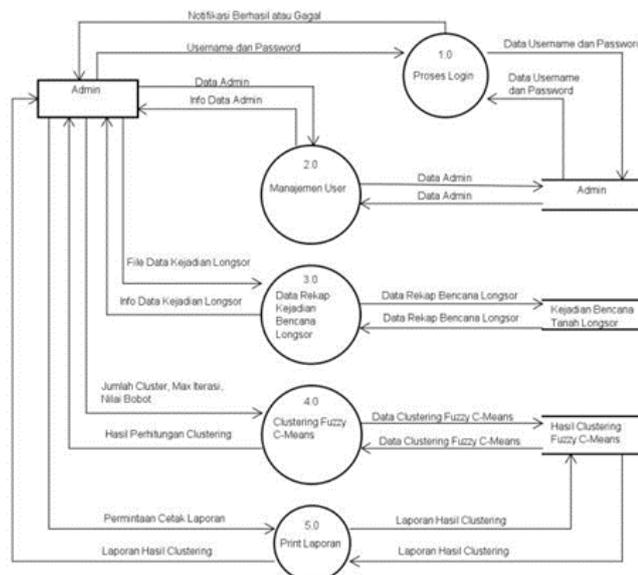
- j. Sistem menentukan nilai bilangan random secara acak dengan nilai keanggotaan pada interval 0 sampai dengan 1.
 - k. Sistem menghitung fungsi objektif pada proses iterasi.
 - l. Jika proses melebihi maksimum iterasi maka proses *fuzzy c-means* berhenti.
 - m. Kemudian admin dapat menyimpan hasil clustering fuzzy c-means.
 - n. Setelah data diproses, program akan menghasilkan nilai *Cluster* berdasarkan klaster yang ditentukan, atribut kecamatan dan tahun kejadian.
 - o. Klik tombol 'Cetak Laporan' untuk mencetak hasil *clustering*.
- 8) Data Flow Diagram (DFD) (Li & Chen, 2009)

a. Diagram Konteks



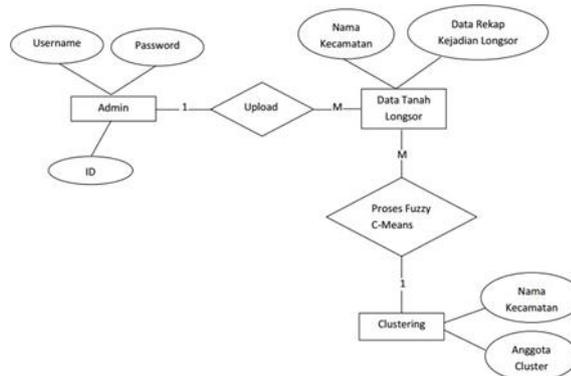
Gambar 5 Diagram Konteks

b. DFD Level 0



Gambar 6 DFD Level 0

c. *Entity Relationship Diagram*



Gambar 7 Entity Relationship Diagram

Sumber: (Yeh et al., 2008)

Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan sesuai dengan skenario pengujian terhadap sistem yang dibangun, maka dapat disimpulkan bahwa: 1). Sistem ini dapat mempermudah instansi BPBD Kabupaten Kuningan dalam melakukan analisis data kebencanaan khususnya bencana longsor berdasarkan wilayah kecamatan dengan lebih terstruktur. 2). Aplikasi ini berhasil melakukan *clustering* atau pengelompokan dengan teknik Fuzzy C-Means yang mana menghasilkan pengelompokan data bencana longsor berdasarkan kecamatan dalam pemanfaatannya kepada instansi BPBD Kabupaten Kuningan dalam rangka melakukan analisis lanjutan. 3). Proses *clustering* dalam aplikasi ini dapat digunakan oleh pihak BPBD Kabupaten Kuningan dalam rangka melaksanakan proses mitigasi bencana.

BIBLIOGRAFI

- Alfina, T., Santosa, B., & Barakbah, A. R. (2012). Analisa perbandingan metode hierarchical clustering, k-means dan gabungan keduanya dalam cluster data (studi kasus: Problem kerja praktek teknik industri its). *Jurnal Teknik Its*, 1(1), A521–A525.
- Arista, R. R., Asmara, R. A., & Puspitasari, D. (2017). Pengelompokan Kejadian Gempa Bumi menggunakan Fuzzy C-Means Clustering. *Jurnal Teknologi Informasi Dan Terapan (J-TIT)*, 4(2), 103–110.
- Bifawaidati, I. (2017). *Penerapan Algoritma Fuzzy C-Means Untuk Klasifikasi Citra Tenun Berdasarkan Fitur Tekstur*. Institut Teknologi Sepuluh Nopember.
- Kemala, R. F., Astuti, I. F., & Maharani, S. (2019). Penerapan Metode Fuzzy C-Means Pada Aplikasi Simulasi Toefl (Test Of English As A Foreign Language) Berbasis Web (Studi Kasus: Fakultas Mipa Universitas Mulawarman). *Jurnal Ilmiah Ilmu Komputer*, 14(1), 17.
- Li, Q., & Chen, Y.-L. (2009). Data flow diagram. In *Modeling and Analysis of Enterprise and Information Systems* (pp. 85–97). Springer.
- Melilisa, I. (2011). *analisis data pembayaran kredit nasabah bank menggunakan metode data mining*. Universitas Multimedia Nusantara.
- Pamusti, R. (2018). *Studi Kinerja Algoritma Optimasi pada Metode Quantum Clustering dengan Kernel Entropy Component Analysis untuk Reduksi Dimensi*. Institut Teknologi Sepuluh Nopember.
- Roger, P., & Pressman, P. D. (2012). *Rekayasa Perangkat Lunak*. Yogyakarta: Andi.
- Sugiyono, S. (2016). *Metode penelitian kuantitatif dan kualitatif dan R&D*. Alfabeta Bandung.
- Syahputra, T., Halim, J., & Sintho, E. P. (2018). Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi Sma Menggunakan Metode Clustering Dengan Teknik Single Linkage. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 4(2), 205–208.
- Tan. (2017). Clustering. Retrieved from Bina Nusantara School of Computer Science. <https://socs.binus.ac.id/2017/03/09/clustering/>
- Yeh, D., Li, Y., & Chu, W. (2008). Extracting entity-relationship diagram from a table-based legacy database. *Journal of Systems and Software*, 81(5), 764–771.