PENERAPAN K- MEDOIDS DALAM KLASIFIKASI PERSEBARAN LAHAN KRITIS DI JAWA BARAT BERDASARKAN KABUPATEN/KOTA

Lita Karlina, Odi Nurdiawan
Program Studi Manajemen Informatika D3, Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon, Indonesia
Litakarlina05@gmail.com

ABSTRAK

Karakteristik lahan kritis meliputi kerusakan struktur tanah, penurunan kuantitas dan kualitas bahan organik, kekurangan unsur hara, dan terganggunya siklus hidrologi. Lahan kritis membutuhkan rehabilitasi dan peningkatan produktivitas untuk kembali ke keadaan sebelumnya sebagai ekosistem yang sehat atau untuk menghasilkan hasil yang lebih baik. Perambahan hutan, penebangan liar, kebakaran hutan, dan penggunaan sumber daya hutan yang tidak tepat adalah semua faktor yang berkontribusi pada lahan penting. Tujuan penelitian ini adalah untuk mengkategorikan alokasi lahan kritis Jawa Barat berdasarkan kabupaten dan kota. Dengan menganalisis berbagai sumber data dan menggunakan alat untuk identifikasi pola seperti statistik dan matematika, data mining adalah proses menemukan koneksi dan pola baru. Peneliti menggunakan Algoritma *K-Medoids*, teknik pengelompokan yang terkait dengan algoritma k-means dan algoritma medoidshift, untuk memecahkan masalah ini. Algoritma *Partitoning Around Medoids* (PAM) adalah nama lain untuk algoritma *k-medoids*. Data yang digunakan dalam penelitian ini berasal dari opendatajabar.com dan mencakup tahun 2014 hingga 2020. Sebanyak 81 item di *cluster* 0, 48 item di *cluster* 1, 229 item di *cluster* 2, 14 item di *cluster* 3, dan 6 item di *cluster* 4, karena data akan diolah menggunakan 5 *cluster*.

Kata Kunci: *Lahan kritis, Data mining, Algoritma K-Medoids*

1. PENDAHULUAN

Berbagai informasi yang akhir - akhir ini muncul di Indonesia, khususnya banjir dan kekeringan, menunjukkan bahwa daya dukung atau daya tampung sumber daya lahan dan lingkungan semakin berkurang [1]. Lahan kritis digambarkan sebagai lahan yang telah mengalami degradasi fisik, kimia, dan biologi akibat tidak memenuhi peruntukan dan kemampuannya, membahayakan hasil pertanian, permukiman, kehidupan sosial ekonomi, dan lingkungan. Kerusakan struktur tanah, penurunan kuantitas dan kualitas bahan organik, kekurangan unsur hara, dan gangguan pada siklus hidrologi adalah tanda-tanda lahan kritis. Kondisi ini menuntut peningkatan produktivitas untuk mengembalikan kemampuan lahan untuk berfungsi sebagai ekosistem yang sehat atau untuk menghasilkan barang yang lebih baik [2].

Untuk menghentikan jumlah lahan kritis berkembang, diperlukan penelitian lebih lanjut tentang masalah ini. Mengingat banyaknya data yang tersedia, informasi ini dapat dikumpulkan dengan menggunakan pendekatan data mining. Teknik pengelompokan data mining sangat membantu untuk mengkategorikan data status lahan berdasarkan kabupaten/kota sehingga dapat dimanfaatkan sebagai landasan untuk memproyeksikan luas lahan kritis di masa depan. Hal ini berlaku untuk data terkait luas lahan kritis. Hasilnya, diharapkan penelitian ini akan mampu mengklasifikasikan data luas lahan penting dengan menggunakan teknik *Knowledge Discovery in Database Process* (KDD) dan algoritma k-medoids berdasarkan kondisi lahan yang ada [3].

Mahasiswa asal STIKOM Tunas dan Putra Pratama Siregar melakukan penelitian. Pada tanggal 4 Maret 2022, dengan judul "Penerapan Metode K-Means dalam Pengelompokan Lahan Kritis di Indonesia berdasarkan Provinsi." Penurunan penggunaan lahan yang tidak mampu lagi untuk mempertahankan pengelolaan air dan sumber daya lahan adalah hasil dari lahan kritis. Pertumbuhan dan perkembangan tanaman pangan, pertanian, perkebunan, kehutanan, perikanan, kota, industri, dan pariwisata juga dianggap tidak mungkin terjadi di medan kritis. Keadaan ini dapat berdampak pada kesejahteraan sosial dan ekonomi mereka yang menggunakan tanah tersebut. Oleh karena itu, untuk menentukan provinsi mana yang memiliki jumlah distribusi lahan penting yang tinggi diperlukan bagi kelompok tersebut. Metode algoritma *K-Means* Data Mining dapat digunakan sebagai solusi untuk masalah ini. Dengan demikian, lahan kritis di Indonesia dapat diklasifikasikan berdasarkan provinsi menggunakan data mining dan teknik K-means. Hasil metode *K-means* dapat diterapkan data mining cepat dengan validasi nilai yang sama; data penelitian ini berasal dari BPS. Sebanyak 34 provinsi masuk dalam kumpulan data uji coba, dengan dua *cluster* digunakan untuk klaster tinggi yang terdiri dari empat provinsi: Sumatera Utara, Jambi, Jawa Timur, dan Kalimantan Tengah. sekelompok kecil yang terdiri dari 30 provinsi.

Tujuan dari penelitian ini adalah menggunakan teknik data mining *Knowledge Discovery in Database Process* (KDD) untuk mengelompokkan luas lahan yang ada di kabupaten/kota di Provinsi Jawa Barat

menjadi 5 *cluster* berdasarkan latar belakang masalah. Dari situ, akan dilakukan pengelompokan lahan kritis berdasarkan tahun 20014-2020.

2. TINJAUAN PUSTAKA

2.1. Lahan Kritis

Daerah kritis adalah daerah yang mengandung sumber daya alam yang dapat dimanfaatkan oleh masyarakat untuk memajukan ekonomi manusia. Hal ini mengakibatkan pemisahan tanah menjadi dua kategori, lahan prospektif dan lahan krusial. Meskipun lahan esensial kurang produktif dibandingkan lahan potensial, namun masih dapat dimanfaatkan untuk kesejahteraan manusia.

Lahan kritis didefinisikan sebagai lahan yang tidak berfungsi sebagai media produksi untuk menanam tanaman budidaya atau tidak dibudidayakan berdasarkan Undang-Undang Nomor 37 Tahun 2014 tentang Konservasi Tanah dan Air. Karena daerah sekitarnya kering dan tidak dapat digunakan untuk budidaya, lahan kritis tidak produktif dan memiliki kesuburan yang sangat rendah. Menilai lahan kritis dan menemukan area kritis menggunakan data geografis parametrik [2].

Karena aktivitas antropogenik dan pengaruh alam, tanah dapat dikategorikan sebagai kritis. Namun, kerusakan lahan biasanya merupakan produk dari aktivitas manusia; Kerusakan ini terjadi sebagai akibat dari penggunaan lahan yang berada di luar kapasitas nasional, yang menyebabkan kerusakan fisik, kimia, dan biologis. Selain aktivitas manusia, perubahan iklim dan bencana alam juga berdampak pada kerusakan lahan. [4].

2.2. Data Mining

Data mining adalah teknik pemindaian otomatis di ruang memori yang sangat besar untuk informasi yang relevan [5]. Data minig kadang-kadang digambarkan sebagai proses yang mengekstraksi informasi yang relevan dan mengumpulkan pengetahuan dari berbagai database atau gudang data yang cukup besar menggunakan metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesi [6].

Salah satu teknik untuk melakukan data mining dikenal sebagai *Knowledge Discovery in Database Process* (KDD). Menurut Fayyed et al. (1996), KDD adalah proses penggunaan teknik data mining untuk menemukan informasi dan pola penting dalam data, termasuk algoritma untuk melakukannya. Data mining, pra-pengolahan data, transformasi data, penambangan data, dan interpretasi adalah tahapan dari *proses Knowledge Discovery in Database* (KDD) yang dijelaskan Dunham (2003) [7].

2.3. Algoritma K-Medoids

Algoritma *Partitioning Around Medoids* (PAM) adalah nama lain untuk algoritma k-medoids. Karena kedua sistem menggunakan algoritma partisi, algoritma *k-medoid* dan *k-means* sebanding. Algoritma untuk mempartisi data ke dalam banyak *cluster* tanpa struktur hierarkis di antara mereka disebut algoritma partisi. Jika dibandingkan dengan metode *k-means*, algoritma *k-medoids* memiliki keunggulan. *K-medoid* berfungsi lebih baik ketika hanya sedikit data yang digunakan. Dalam algoritma ini, *cluster* diwakili oleh objek dalam kumpulan objek. [8] *Medoids* adalah item yang dipilih untuk menggambarkan sebuah *cluster*. Perhitungan algoritma k-medoids meliputi, antara lain:

a. Inisialisasi pusat kluster (jumlah *cluster*)
b. Gunakan persamaan pengukuran jarak Jarak Euclidian untuk menetapkan setiap bagian data (objek) ke *cluster* terdekat.

$$d_{ij} = \sqrt{\sum_{a=1}^{p} \left(X_{ia} - X_{ja} \right)^2} = \sqrt{\left(X_i - X_j \right)' \left(X_i - X_j \right)}$$

di mana p dan V adalah matriks varian kovarians, dan I, j, dan n semuanya bilangan bulat.
c. Pilih kandidat acak untuk *medoids* baru dari setiap kelompok item.
d. Tentukan pemisahan antara setiap objek di setiap *cluster* menggunakan kandidat medoid baru.
e. Tentukan deviasi total (S) dengan membagi total jarak sebelumnya dengan nilai total jarak baru. Untuk membuat koleksi baru objek k yang *medoids*, tukar objek dengan kluster data jika S 0.
f. Terus lakukan langkah 3 sampai 5 sampai tidak ada lagi modifikasi pada medoid, di mana titik cluster dan masing-masing anggota cluster diperoleh.

2.4. Rapidminer

Menyelesaikan pekerjaan analisis canggih dengan mudah menyelesaikan pekerjaan analisis canggih dengan mudah bahkan tanpa banyak keterampilan pemrograman. Antarmuka pengguna *RapidMiner* mudah digunakan, sehingga memudahkan pengguna untuk membangun alur kerja dan melihat hasil [9].

Fitur-fitur berikut membedakan RapidMiner dari perangkat lunak analisis data lainnya:

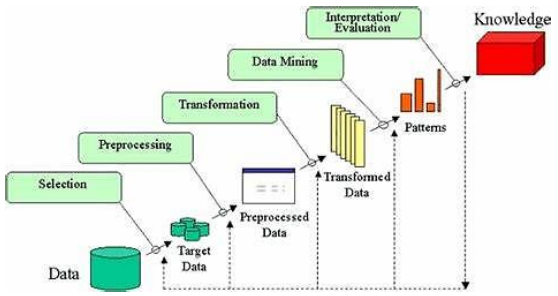
a. Pengguna dapat menentukan alur kerja dan mengonfigurasi proses tanpa menulis kode berkat antarmuka drag-and-drop.
b. Beberapa operator *RapidMiner* memiliki ribuan operator yang tersedia untuk digunakan dalam membersihkan, mengonversi, dan menganalisis data, di antara tugas analitis lainnya.
c. Banyak jenis data *RapidMiner* dapat memproses berbagai jenis data, termasuk teks, gambar, bahasa, dan data tabular.
d. Kemampuan pembelajaran mesin *Rapid Miner* mencakup berbagai algoritma pembelajaran mesin, termasuk pengelompokan, regresi, klasifikasi, dan visualisasi.
e. Adaptasi ke sumber data yang berbeda Pengguna *RapidMiner* dapat menggabungkan data dari berbagai sumber, termasuk file, database, dan API.

f. Ada beberapa visualisasi yang tersedia di *RapidMiner* yang dapat digunakan untuk menyajikan data dan analisis dengan cara yang mudah dipahami.

3. METODE PENELITIAN

3.1. Perancangan Penelitian

Metode *Knowledge Discovery in Database Process* (KDD) digunakan dalam desain penelitian. Fayyad et al. (1996) mendefinisikan *Knowledge Discovery in Database Process* (KDD) sebagai proses dimana metode data mining digunakan untuk mencari informasi dan pola yang relevan dalam data, sedangkan data mining menggunakan algoritma untuk menemukan pola untuk mengidentifikasi dalam suatu proses database. Penambangan data adalah salah satu proses *end-to-end* yang mencakup penemuan pengetahuan dalam database (KDD) [9]. Berikut ini adalah beberapa langkah dalam proses *Knowledge Discovery in Database Process* (KDD):



Gambar 1. Proses Knowledge Discovery in Database Process (KDD)

1. Data Selection
   Sebelum beralih ke langkah penambangan informasi dari proses *Knowledge Discovery in Database Process* (KDD), data harus dipilih dari sekumpulan data operasional. Data yang dipilih proses penambangan data disimpan dalam file terpisah dari database operasi.
2. Pre-Processing/Cleaning
   Knowledge Discovery in Proses Database (KDD) berfokus pada pembersihan data sebelum proses data mining dapat dilakukan. Duplikasi data harus dihilangkan, data yang tidak konsisten harus diperiksa, dan ketidakakuratan dalam data harus diperbaiki, antara lain, selama proses pembersihan.
3. Transformasi
   Sebelum dapat digunakan, beberapa teknik penambangan data memerlukan format data tertentu. Kualitas hasil data mining juga dipengaruhi oleh modifikasi dan seleksi data ini. Pada tingkat ini, beberapa pendekatan penambangan data tertentu bergantung.
4. Data Mining
   Data mining adalah praktik menggunakan alat atau prosedur tertentu untuk mengungkap pola atau informasi menarik dalam satu set data. Metodologi, teknik, dan algoritma penambangan data bisa sangat berbeda. Pendekatan atau algoritma terbaik untuk digunakan sangat ditentukan oleh tujuan keseluruhan dan Penemuan Pengetahuan dalam Proses Basis Data (KDD).
5. Interpretasi dan penilaian
   Penting untuk menyajikan pola informasi proses penambangan data dengan cara yang dapat dimengerti oleh pihak yang berkepentingan. Fase ini termasuk dalam kategori interpretasi dalam proses *Knowledge Discovery in Database Process* (KDD). Memeriksa apakah pola atau data yang ditemukan bertentangan dengan fakta atau hipotesis sebelumnya adalah bagian dari tahap ini. Teknik untuk memuat tesis dijelaskan di bagian ini.

3.2. Pengumpulan Data

Data luas lahan vital berdasarkan status lahan di Provinsi Jawa Barat tahun 2014 hingga 2020, dengan total data sebanyak 378, digunakan dalam penelitian ini. https://opendata.jabarprov.go.id/id.

4. HASIL DAN PEMBAHASAN

Langkah-langkah berikut digunakan untuk memodelkan pada tahap ini menggunakan Rapidminer versi 10.1:

4.1. Data Selection

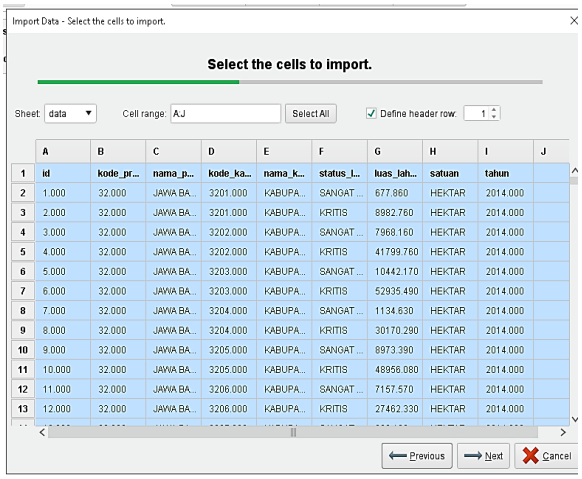
Untuk menghilangkan data yang tidak bertentangan atau duplikat dan memperbaiki kesalahan pada data yang ada, data dimasukkan ke dalam proses seleksi. Setelah pemilihan data, 378 data, mirip dengan data yang ditunjukkan pada table 1, akan digunakan sebagai berikut:

Tabel 1. Dataset Luas Lahan Kritis di Jawa Barat

id	Kode Kabupaten Kota	Nama kabupaten kota	Status Lahan	Luas lahan kritis	Tahun
1	3201	KABUPATEN BOGOR	SANGAT KRITIS	677.86	2014
2	3201	KABUPATEN BOGOR	KRITIS	8982.76	2014
3	3202	KABUPATEN SUKABUMI	SANGAT KRITIS	7968.16	2014
4	3202	KABUPATEN SUKABUMI	KRITIS	41799.76	2014
5	3203	KABUPATEN CIANJUR	SANGAT KRITIS	10442.17	2014
6	3203	KABUPATEN CIANJUR	KRITIS	41799.76	2014
……	……..	……..	……..	…….	….
376	3278	KOTA TASIKMALAYA	KRITIS	178.07	2020
377	3279	KOTA BANJAR	SANGAT KRITIS	21.46	2020
378	3279	KOTA BANJAR	KRITIS	598.54	2020

4.2. Pre-Processing/Cleaning

Mengikuti pemilihan data dan atribut yang akan digunakan, pra-pemrosesan data dilakukan untuk menghilangkan data duplikat, mengisi celah dalam data, dan memperbaiki kesalahan yang mungkin telah terjadi. Untuk mengolah data dan melakukan proses data mining, akan dilakukan pembersihan atau pembersihan data pada saat ini.

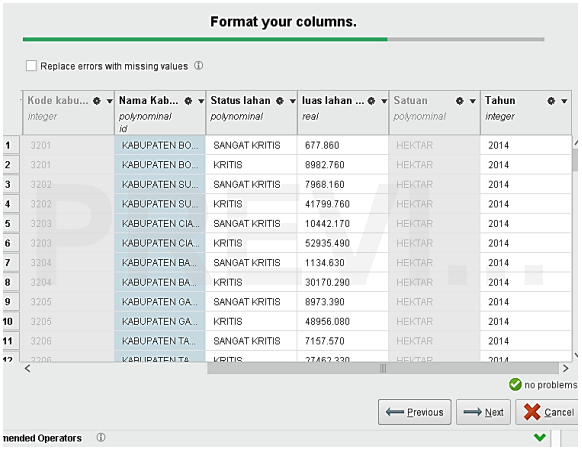


Gambar 2. Pre-Processing Data

4.3. Transformasi Data

Pada titik ini, data dimodifikasi dengan cara yang konsisten dengan proses penambangan data. Microsoft Excel digunakan untuk membangun himpunan data yang akan dimuat ke Rapidminer, yang berisi file yang akan digunakan dalam penelitian ini. Tentukan masing-masing indikasi atribut dalam file setelah diimpor. Jenis atribut adalah polinominal karena semua atribut memiliki lebih dari dua kelas.

Selama impor data, properti nama distrik dimodifikasi atau dipilih ke (id). Nama wilayah atau wilayah memiliki kuncinya sendiri untuk mengidentifikasi nama wilayah, itulah sebabnya ia dimodifikasi.

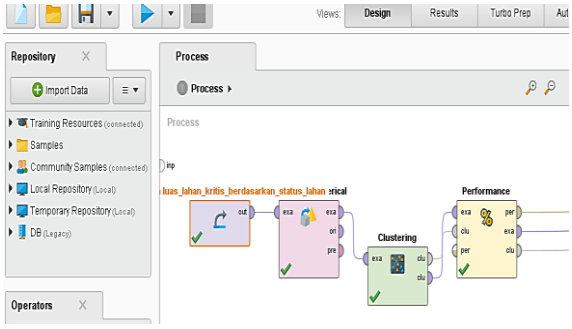


Gambar 3. Transformasi Data

4.4. Data Mining

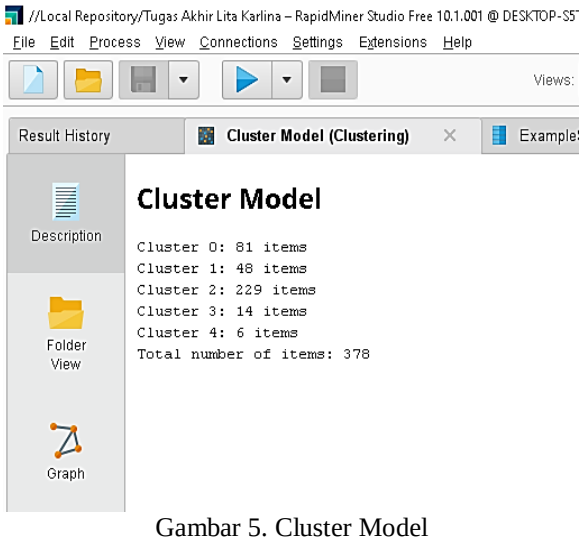
Tahapan data mining pada tugas akhir ini menggunakan metode *clustering K-Medoids, u*ntuk membantu melakukan proses penggalian data dengan langkah sebagai berikut:

a. Pilih Import data lalu masukan data excel yang akan di uji.
b. Setelah mengimport data excel lalu drag ke panel process.
c. Pada panel operators ketik Nominal to Numerical lalu drag ke panel process.
d. Ketik K-Medoids di panel operator, seret ke panel proses, dan tentukan berapa banyak K yang harus digunakan.
e. Ketik "Performance" di panel operator dan seret ke panel proses. Operator parameter kriteria utama dimodifikasi ke Davies Bouldin, dan fungsi normalisasi dan maksimalkan pemeriksaan ditambahkan.
f. Untuk melihat hasil pengelompokan, klik tombol Run setelah menghubungkan konektor ke setiap proses di panel proses.



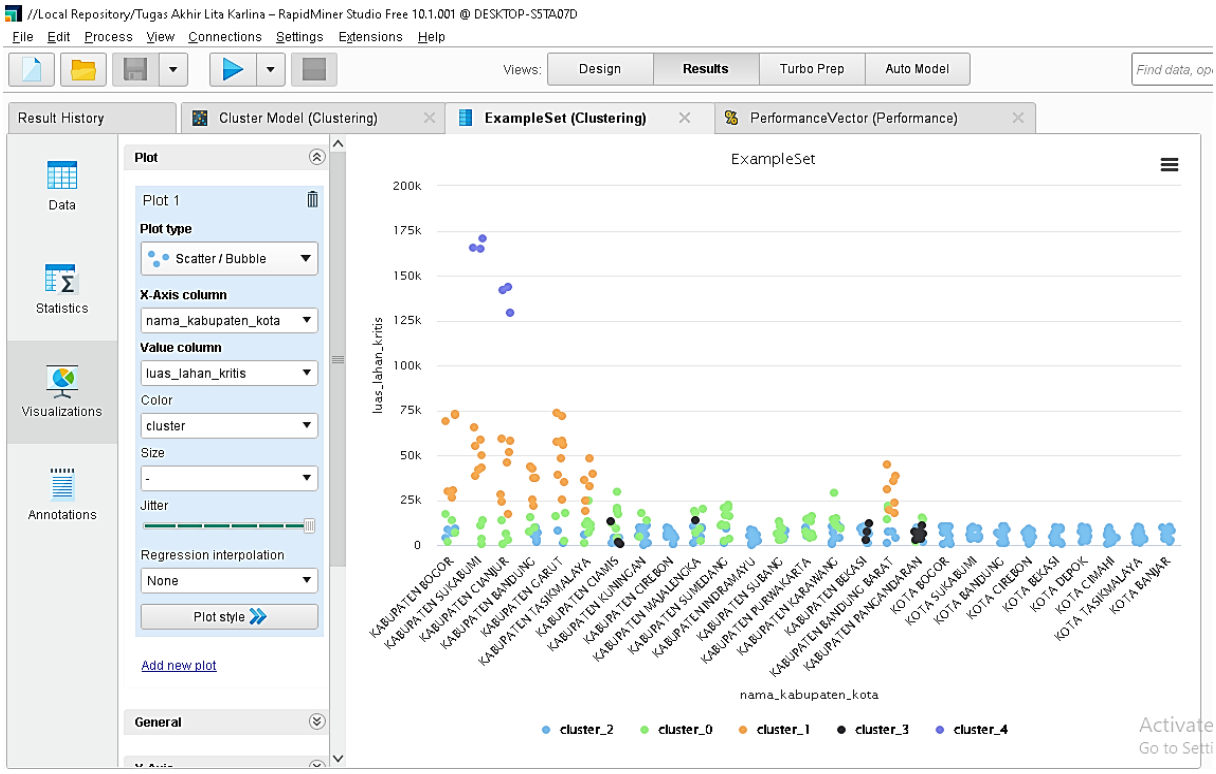
Gambar 4. Data Mining

4.5. Interpretation/Evaluation



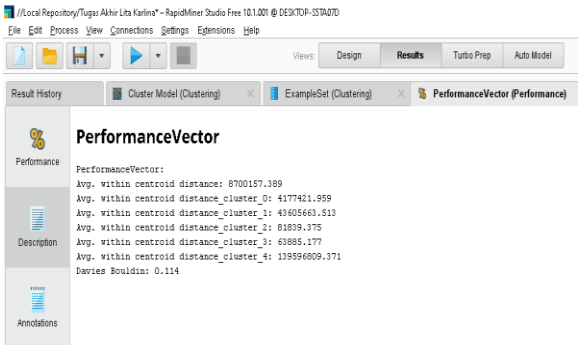
Gambar 5. Cluster Model

81 item di cluster 0, 48 item di cluster 1, 229 item di cluster 2, 14 item di cluster 3, dan 6 item di cluster 4, menghasilkan total 5 cluster.



Gambar 6. Hasil Visualisasi Algoritma K-Medoids

Total dataset yang dikelompokkan sebanyak 378 data, dengan jumlah wilayah di setiap cluster, dan setelah proses data mining selesai, maka akan muncul hasil dataset dengan menggunakan metode yang dipilih dan menghasilkan 5 cluster mulai dari cluster 0, cluster 1, cluster 2, cluster 3, dan cluster 4. Hasil model cluster dan visualisasi dapat dilihat pada gambar 5 dan 6, dan pada perhitungan ini menghasilkan Davies Bouldin Indeks seperti pada gambar 7 berikut:



Gambar 7. Nilai Davies Bouldin Indeks

Dari hasil pengklusteran cluster 5 maka menghasilkan nilai davies bouldin indeks sebesar 0.114, rata-rata pada jarak pusat: 8700157.389, rata-rata dalam jarak pusat 0: 4177421.959, rata-rata pada jarak pusat 1: 43605663.513, rata-rata pada jarak pusat 2: 81839.375, rata-rata pada jarak pusat 3: 63885.177, dan rata-rata pada jarak pusat 4: 139596809.371.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil tugas akhir dan pembahasan, maka diperoleh kesimpulan sebagai berikut: Dari hasil clustering yang terbaik, maka diperoleh 5 cluster yaitu 81 wilayah pada cluster 0, 48 wilayah pada cluster 1, 229 wilayah pada cluster 2, 14 wilayah pada cluster 3, 6 wilayah pada cluster 4. Dari pencarian Nilai Davies Bouldin Indeks hasil mengelompokan luas lahan kritis menggunakan algoritma *k-medoids* clustering dengan menggunakan 5 cluster maka menghasilkan nilai davies bouldin indeks sebesar 0.114, rata-rata dalam jarak pusat: 8700157.389, rata-rata dalam jarak pusat 0: 4177421.959, rata-rata dalam jarak pusat 1: 43605663.513, rata-rata dalam jarak pusat 2: 81839.375, rata-rata dalam jarak pusat 3: 63885.177, rata-rata dalam jarak pusat 4: 139596809.371.

Tentu saja, ada keterbatasan dalam penelitian ini, sehingga para ahli menawarkan rekomendasi untuk penelitian di masa depan untuk memperbaikinya lebih baik. Saran peneliti adalah sebagai berikut: Harus dilakukannya proses text preprocessing yang jauh lebih baik lagi supaya tidak ada data yang duplikat. Untuk membandingkan algoritma pengelompokan dan klasifikasi lain dan mencapai hasil yang lebih tepat atau unggul, studi tambahan harus dilakukan di bidang ini.

DAFTAR PUSTAKA

[1] U. Kurnia, N. Sutrisno, and I. Sungkana, "Perkembangan Lahan Kritis," *Balai Besar Litbang Sumber Daya Lahan Pertan.*, pp. 144–160, 2010.

[2] N. Bashit, "Analisis Lahan Kritis Berdasarkan

Kerapatan Tajuk Pohon Menggunakan Citra Sentinel 2," *Elipsoida J. Geod. dan Geomatika*, vol. 2, no. 01, pp. 71–79, 2019, doi: 10.14710/elipsoida.2019.5019.

[3] Putra Pratama Siregar, S Solikhun, and Zulia Almaida Siregar, "Penerapan Metode K-Means Dalam Mengelompokkan Persebaran Lahan Kritis Di Indonesia Berdasarkan Provinsi," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 4, pp. 145–151, 2022, doi: 10.30865/resolusi.v2i4.335.

[4] Rimba kita, "Lahan Kritis-Pengertian, Penyebab, ciri, Data Sebaran & Solusi," *Website*, 2019. https://rimbakita.com/lahan-kritis/

[5] 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.

[6] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.

[7] W. Latuny, "Memprediksi Harga Jual Rumput Laut Kering Pada Tingkat Petani Dengan Data Mining," *ALE Proceeding*, vol. 2, no. April, pp. 186–199, 2021, doi: 10.30598/ale.2.2019.186-199.

[8] S. R. Ningsih, I. S. Damanik, A. P. Windarto, and H. Satria, "Analisis K-Medoids Dalam Pengelompokkan Penduduk Buta Huruf Menurut Provinsi," no. September, pp. 721–730, 2019.

[9] R. Fikri, Nisa, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Alfabeta, 2020.