

Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC

Mitha Amelia Sundari, Rahmadhani Pane*, Rohani

Prodi Teknologi Informasi, Fakultas: Sains dan Teknologi, Universitas Labuhanbatu, Indonesia

Email: ¹ameliamitha78@gmail.com, ²rahmadanipane@gmail.com, ³pasariburohani@gmail.com

Email Penulis Korespondensi: ²rahmadanipane@gmail.com

Submitted: 22/05/2023; Accepted: 29/06/2023; Published: 29/06/2023

Abstrak—Pelecehan seksual merupakan salah satu kejahatan kriminal yang sering terjadi di Indonesia belakangan ini. Tindakan pelecehan seksual dapat terjadi dalam kehidupan sehari-hari tanpa memandang waktu, baik di tempat kerja, di jalan, maupun di rumah sendiri. Wanita sering menjadi korban pelecehan seksual, meskipun pria juga dapat mengalami hal serupa. Pelaku pelecehan seksual dapat berasal dari orang yang tidak dikenal, orang yang memiliki kebencian, bahkan orang yang kita sayangi sekalipun. Kurangnya pendidikan agama, moral, dan perkembangan teknologi yang memungkinkan akses mudah terhadap konten pornografi menjadi faktor penyebab terjadinya pelecehan seksual. Untuk mengatasi masalah ini, diperlukan tindakan cepat di tempat-tempat yang sering terjadi pelecehan seksual melalui sosialisasi agar masyarakat lebih waspada saat berada di tempat tersebut. Selain itu, perlu ditingkatkan keamanan di wilayah tersebut dan disediakan tempat konsultasi seperti psikolog. Untuk mengidentifikasi tempat-tempat yang rawan terjadinya pelecehan seksual di Indonesia, diterapkan metode data mining dengan memanfaatkan data-data sebelumnya. Metode pengelompokan yang digunakan adalah AHC dengan menggunakan mode complete linkage (jarak terjauh) antara kluster awal. Hasil akhir dari penelitian ini melibatkan proses manual dan aplikasi RapidMiner yang sesuai, sehingga dapat melanjutkan pembentukan kluster baru menggunakan RapidMiner. Terdapat 5 provinsi yang termasuk dalam cluster 0, kemudian terdapat 17 provinsi dalam cluster 1, dan 12 provinsi dalam cluster 2.

Kata Kunci: Data Mining; Clusterieng; AHC; Complate Linkage; Korban Kejahatan Pelecehan Seksual

Abstract—Sexual harassment is one of the most common crimes in Indonesia recently. Acts of sexual harassment can occur in everyday life regardless of time, whether at work, on the street, or at home. Women are often the victims of sexual harassment, although men can experience the same. Perpetrators of sexual harassment can come from people we don't know, people who have hatred, even people we care about. Lack of religious and moral education, and technological developments that allow easy access to pornographic content are contributing factors to sexual harassment. To overcome this problem, fast action is needed in places where sexual harassment often occurs through socialization so that people are more vigilant when they are in these places. Apart from that, it is necessary to improve security in the area and provide consultation places such as psychologists. To identify places that are prone to sexual harassment in Indonesia, a data mining method is applied by utilizing previous data. The clustering method used is AHC using the complete linkage mode (longest distance) between the initial clusters. The final results of this research involve a manual process and the appropriate RapidMiner application, so that new clusters can be formed using RapidMiner. There are 5 provinces included in cluster 0, then there are 17 provinces in cluster 1, and 12 provinces in cluster 2.

Keywords: Data Mining; Clustering; AHC; Complete Linkage; Victims of Sexual Harassment Crimes

1. PENDAHULUAN

Kejahatan merupakan sebuah tindakan negatif yang tidak lepas dari kehidupan seseorang, tidak peduli tempat dan waktu keberadaan seseorang, kejahatan dapat terjadi begitu saja, selain dapat terjadi kapan saja dan dimana saja, kejahatan juga tentunya berdampak buruk baik bagi sipelaku kejahatan ataupun korban dari kejahatan tersebut dan juga dapat merugikan diri sendiri ataupun orang lain. Tindak kejahatan dikenal juga sebagai kriminal yang melanggar hukum baik berupa norma-norma ataupun peraturan yang telah ditetapkan sehingga seseorang yang melakukan tindak kriminal harus diberi hukuman. Di Indonesia ada banyak tindak kriminal yang sering terjadi dengan berbagai kasus seperti perampokan (pencurian), pembunuhan, pelecehan seksual, penganiayaan, dan banyak lagi jenis kriminal yang terjadi di Indonesia.

Pada penelitian ini, kasus yang akan dibahas yaitu tingkat pelecehan seksual yang terjadi di Indonesia meningkat dengan persentase 1,24 pada tahun 2020 menjadi 1,71 pada tahun 2021. Pelecehan seksual hampir dapat ditemui dalam kehidupan sehari-hari tanpa pandang waktu. Baik di lingkungan kerja, di jalan, bahkan di rumah sendiri pun kerap terjadi pelecehan seksual. Pelecehan seksual sering terjadi pada wanita dan juga bahkan terjadi juga bagi pria, hanya saja wanita masih lebih sering menjadi target pelecehan seksual daripada pria. Selain tempat, waktu dan target terjadi pelecehan seksual, pelaku pelecehan seksual dapat berupa orang yang tidak dikenal, yang memiliki kebencian dan bahkan orang yang kita anggap sangat menyayangi kita juga dapat melakukan tindakan tercela tersebut. Semua itu disebabkan karena kurangnya didikan dari orang tua terhadap agama, moral dan perkembangan teknologi yang tidak lagi membatasi pengguna untuk mengakses hal-hal yang bersifat negatif seperti mengakses konten pornografi, Untuk mengatasi hal tersebut maka diperlukan sebuah tindakan yang cepat terhadap tempat yang dominan terjadinya pelecehan seksual untuk dilakukan sosialisasi agar lebih waspada jika berada di tempat (kota) tersebut, meningkatkan keamanan di wilayah tersebut dan juga menyediakan tempat konsultasi seperti seorang psikolog.

Berdasarkan masalah diatas, maka perlu diterapkan data mining untuk mengetahui tempat yang rawan terjadi korban pelecehan seksual di Indonesia berdasarkan data-data yang ada sebelumnya, dalam data mining, beberapa kelompok yang dapat diterapkan dengan metode penggunaan masing-masing diantaranya melakukan prediksi terhadap objek dengan menggunakan metode SVM, Naïve Bayes, C4.5, Linear Regression, Knn dan Rought Set [1],

[2]. Melakukan pengelompokkan dengan memanfaatkan metode clustering seperti K-Medoids, K-Means, AHC, SOM, dan lainnya [3], [4]. Selin itu dapat juga melakukan Pengklasifikasian dengan menggunakan metode Cart, C4.5, ID3, K-NN, Naive Bayes dan lainnya [5], [6]. Ada juga kelompok data mining lainnya yang dapat digunakan seperti estimasi dan asosiasi. Pada kasus penelitian ini akan dilakukan pengelompokkan korban pelecehan seksual di Indonesia dengan menggunakan metode AHC Clustering. Metode AHC dalam membentuk cluster dapat menggunakan proses single linkage, complete linkage dan Average Linkage [7]–[9].

Ada beberapa penelitian terdahulu yang telah dilakukan diantaranya Rima Tamara Aldisa yang melakukan penelitian pada tahun 2023, dalam penelitiannya tersebut bahwa banyak siswa menghadapi kesulitan dalam proses belajar karena mereka memilih jurusan yang tidak sesuai dengan kemampuan mereka, dipengaruhi oleh orang lain. Guru juga menghadapi kesulitan dalam menyesuaikan kemampuan setiap siswa secara individual. Untuk mengatasi masalah ini, dilakukan pengelompokkan siswa berdasarkan tingkat pengetahuan yang sama menggunakan teknik data mining. Penggunaan metode Agglomerative Hierarchical Clustering (AHC) dengan metode single link (single linkage method) terbukti efektif karena dapat mengelompokkan siswa berdasarkan kriteria jarak terdekat dari semua alternatif. Dalam penelitian ini, tiga kriteria digunakan (Nilai IPA, Nilai IPS, Nilai TPA) dengan jumlah siswa sebanyak 121 orang. Penerapan metode AHC menggunakan software RapidMiner untuk efisiensi dan efektivitas. Hasilnya, terdapat empat kluster yang dapat digunakan oleh pihak sekolah: kluster 0 dengan 93 siswa, kluster 1 dengan 10 siswa, kluster 2 dengan 10 siswa, dan kluster 3 dengan 8 siswa [10].

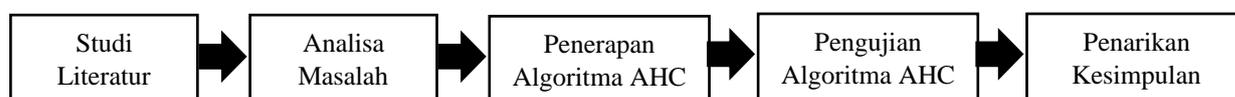
Penelitian selanjutnya dilakukan oleh Raisya Rahma dan Ratna Mufidah pada tahun 2022 bahwa kekerasan dapat menyebabkan dampak negatif baik secara disengaja maupun tidak disengaja terhadap fisik dan psikis seseorang. Perempuan dan anak sering kali menjadi korban kekerasan. Menurut rilis tahunan Komisi Nasional Anti Kekerasan terhadap Perempuan (Komnas Perempuan) tahun 2020, Provinsi Jawa Barat menempati posisi tertinggi kedua di Indonesia dalam kasus kekerasan terhadap perempuan dan anak. Untuk membantu pemerintah dalam menekan angka kasus kekerasan tersebut, dilakukan pemetaan daerah rawan kekerasan. Penelitian ini menggunakan algoritma k-means untuk mengelompokkan daerah rawan kekerasan berdasarkan data Dinas Pemberdayaan, Perempuan, Perlindungan Anak dan Keluarga Berencana tahun 2020-2021. Hasil pengelompokkan membagi daerah menjadi tiga kelompok (cluster): cluster 0 dengan 5 kabupaten atau kota, cluster 1 dengan 21 kabupaten/kota, dan cluster 2 dengan 1 kabupaten. Evaluasi pengelompokkan menggunakan silhouette coefficient menghasilkan nilai sebesar 0,62, menunjukkan bahwa kriteria pengelompokkan yang dilakukan termasuk dalam struktur cluster yang standar (medium structure) [11].

Pada tahun 2023 Lisna Zahrotun, dkk melakukan penelitian tentang Dinas Perindustrian Koperasi Usaha Kecil dan Menengah yang memiliki tanggung jawab terhadap pertumbuhan dan kestabilan Usaha Mikro, Kecil, dan Menengah (UMKM) di kota X, terutama UMKM bidang jasa. Karena banyaknya UMKM bidang jasa, diperlukan penanganan khusus untuk menjaga pertumbuhan dan kestabilan mereka. Untuk memberikan strategi yang tepat, Dinas tersebut perlu melakukan analisis dan identifikasi persebaran UMKM bidang jasa. Salah satu teknik yang digunakan adalah pengelompokkan. Hasil penelitian menunjukkan bahwa metode AHC complete linkage dapat diterapkan pada UMKM bidang jasa, dengan nilai pengujian terbaik sebesar 0,729 dan terdapat dua kelompok yang terbentuk. Kelompok 1 merupakan kelompok UMKM bidang jasa yang perlu mendapatkan perhatian khusus dari Dinas Perindustrian dan Koperasi karena meskipun telah berdiri lama, omsetnya masih di bawah 10 juta [12]. Penelitian juga dilakukan oleh Sinta Maulina Dewi, dkk pada tahun 2019 yaitu sebuah penelitian yang membahas tentang analisis metode K-means dalam pengelompokkan kriminalitas menurut provinsi dimana kejahatan saat ini menjadi perhatian utama dan berita yang sering muncul di berbagai media massa dan elektronik. Salah satu aksi kejahatan yang paling umum adalah pembunuhan, baik yang direncanakan maupun tidak. Aksi ini menimbulkan kekhawatiran di kalangan masyarakat karena banyaknya korban yang ada di sekitar mereka. Hasil dari analisis metode K-means pada pengelompokkan kriminalitas menurut provinsi menunjukkan bahwa terdapat dua kluster yang terbentuk. Kluster tingkat tinggi (C1) terdiri dari 8 provinsi, antara lain Sumatera Utara, Sumatera Selatan, Metro Jaya, Jawa Barat, Jawa Timur, Sulawesi Selatan, Sulawesi Tengah, dan Papua. Sedangkan kluster tingkat rendah (C2) terdiri dari 23 provinsi, termasuk Aceh, Sumatera Barat, Riau, Jambi, Bengkulu, dan 18 provinsi lainnya [13].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Gambar 1 berikut merupakan tahapan penelitian yang dapat dilakukan agar proses penyelesaian lebih terstruktur.



Gambar 1. Tahapan Penelitian

Berdasarkan gambar 1, telah dibuat tahapan yang akan dilakukan untuk menyelesaikan penelitian ini, berikut penjelasan dari gambar tersebut:

a. Studi Literatur

Pada tahap awal sebuah penelitian, melakukan studi literatur atau tinjauan pustaka dimana tahapan tersebut menjadi aspek yang sangat penting dalam menjalankan sebuah penelitian yang berkualitas. Tinjauan pustaka dilakukan dengan mengumpulkan, menganalisis, mengevaluasi, dan menyusun sumber-sumber informasi atau literatur dari berbagai jenis sumber seperti buku, jurnal, artikel, laporan, dan dokumen lain yang relevan dengan topik atau permasalahan yang akan diteliti. Tujuan dari tinjauan pustaka adalah untuk memperoleh pemahaman yang lebih mendalam tentang topik atau permasalahan yang sedang diteliti, serta untuk mengidentifikasi area pengetahuan yang masih kurang atau perlu diteliti lebih lanjut.

b. Analisa Masalah

Analisis masalah dilakukan untuk memperoleh pemahaman dan mengidentifikasi akar permasalahan yang sedang dihadapi. Tujuan dari analisis masalah adalah untuk menemukan penyebab utama dari masalah tersebut sehingga solusi yang tepat dan efektif dapat ditemukan. Proses analisis masalah melibatkan pengumpulan data, analisis data, dan interpretasi data guna mengidentifikasi faktor-faktor penyebab masalah serta dampaknya pada sistem yang terkena dampak. Melalui analisis masalah, tujuan utamanya adalah untuk mendapatkan pemahaman yang komprehensif mengenai masalah yang sedang dihadapi dan mengapa masalah tersebut muncul. Hal ini melibatkan identifikasi dan penelusuran penyebab-penyebab utama yang mempengaruhi masalah tersebut. Dengan memahami akar masalah, langkah-langkah yang tepat dan relevan dapat diambil untuk menyelesaikan atau mengatasi masalah yang ada. Analisis masalah berperan penting dalam mengarahkan upaya penyelesaian masalah dengan cara yang tepat dan efektif. Dengan memahami akar penyebab masalah, solusi yang diusulkan dapat diarahkan untuk menangani sumber permasalahan, bukan hanya gejala yang muncul.

c. Penerapan Algoritma AHC

Setelah proses analisis dilakukan, tindakan selanjutnya yang diambil oleh peneliti adalah menerapkan algoritma pada data yang telah terkumpul. Algoritma digunakan untuk memproses data dengan metode yang telah ditentukan. Dalam penelitian ini, metode yang digunakan untuk menyelesaikan masalah adalah pendekatan AHC (Agglomerative Hierarchical Clustering).

d. Pengujian Algoritma AHC

Setelah penerapan algoritma AHC, langkah selanjutnya adalah melakukan evaluasi terhadap kinerja algoritma tersebut. Evaluasi algoritma dilakukan dengan menggunakan platform RapidMiner. Jika hasil evaluasi konsisten dengan hasil implementasi algoritma sebelumnya, maka evaluasi dianggap berhasil.

e. Kesimpulan

Tahap terakhir dari tahapan penelitian yaitu membuat sebuah kesimpulan terhadap penelitian yang dilakukan, kemudian disusun berbentuk rangkuman.

2.2 Data Mining

Data mining adalah proses penemuan pola dan informasi yang berguna dari sekumpulan data yang besar. Tujuan utama dari data mining adalah mengidentifikasi pola-pola tersembunyi, hubungan, dan wawasan yang dapat digunakan untuk membuat keputusan. Data mining dapat digunakan dengan teknik statistik, database, machine learning dan statistik [14]–[19]. Data mining memiliki kelompok penyelesaian berdasarkan permasalahan masing-masing diantaranya [20]–[22]:

a. Klasifikasi

Klasifikasi merupakan proses pengklasifikasian dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan sebelumnya. Tujuan utama dari klasifikasi adalah mengenali pola-pola yang membedakan data ke dalam berbagai kelas berdasarkan atribut yang relevan. Model klasifikasi digunakan untuk memprediksi kelas atau label yang sesuai untuk data yang belum memiliki kelas. Sebagai contoh, dalam mengklasifikasikan pelanggan ke dalam kategori "pengguna reguler" atau "pengguna potensial," model klasifikasi dapat menganalisis atribut seperti usia, pendapatan, dan riwayat pembelian untuk menentukan kategori yang tepat bagi pelanggan baru tersebut. Klasifikasi memiliki beragam aplikasi, seperti pengenalan pola, deteksi spam email, identifikasi penyakit berdasarkan gejala, penentuan profil pelanggan, dan lain sebagainya. Dengan menggunakan teknik klasifikasi, data dapat diklasifikasikan dengan cepat dan efisien, dan prediksi dapat dibuat berdasarkan pola-pola yang ditemukan dalam data yang ada. Metode klasifikasi yang dapat digunakan yaitu Cart, C4.5, ID3, K-NN, Naive Bayes dan lainnya [23]–[26].

b. Clustering

Clustering merupakan metode dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelompok atau klaster berdasarkan karakteristik atau atribut tertentu. Tujuannya adalah untuk mengidentifikasi pola alami atau struktur tersembunyi dalam data tanpa menggunakan kategori atau label sebelumnya. Proses clustering melibatkan penggunaan algoritma yang menganalisis kesamaan atau perbedaan antara titik-titik data. Algoritma ini berusaha untuk mengelompokkan data ke dalam klaster sehingga data dalam satu klaster memiliki tingkat kesamaan

yang tinggi, sementara data antar kluster memiliki perbedaan yang signifikan. Metode yang dapat digunakan adalah K-Means, AHC, K-Medoids, dan lainnya [27], [28].

c. Asosiasi

Asosiasi merupakan konsep dalam data mining yang berkaitan dengan keterkaitan atau hubungan antara item atau atribut dalam suatu dataset. Analisis asosiasi bertujuan untuk mengungkap pola tersembunyi dalam data yang menunjukkan keterkaitan atau kecenderungan tertentu antara item-item tersebut. Proses analisis asosiasi melibatkan pencarian itemset, yaitu himpunan item yang sering muncul bersama dalam transaksi atau kejadian. Metode yang dapat digunakan dalam proses asosiasi adalah apriori, Fp-Growth, dan lainnya [29], [30].

d. Estimasi

Estimasi merupakan proses perhitungan atau perkiraan nilai atau jumlah yang tidak diketahui berdasarkan informasi atau data yang ada. Tujuannya adalah memberikan perkiraan yang mendekati nilai yang sebenarnya, walaupun tidak ada kepastian mutlak. Dalam data mining, estimasi melibatkan penggunaan teknik statistik atau algoritma untuk menghitung atau memprediksi nilai yang tidak diketahui berdasarkan data yang tersedia. Estimasi dapat diterapkan pada berbagai variabel atau parameter, seperti pendapatan, populasi, pengeluaran, dan lain sebagainya. Estimasi juga digunakan secara luas dalam berbagai bidang, seperti ekonomi, keuangan, ilmu sosial, dan ilmu pengetahuan lainnya. Misalnya, estimasi dapat digunakan untuk memperkirakan pertumbuhan pasar, melakukan proyeksi pendapatan masa depan, atau mengukur tingkat kepuasan pelanggan. Adapun algoritma yang dapat digunakan dalam proses estimasi adalah Expectation Maximization, Regresi Linear berganda, Regresi linear sederhana, dan lainnya [31]–[33].

e. Prediksi

Prediksi merupakan proses atau hasil dari melakukan estimasi atau perkiraan terhadap nilai atau kejadian yang akan terjadi di masa depan berdasarkan informasi atau data yang tersedia saat ini. Tujuannya adalah untuk memberikan perkiraan yang masuk akal atau memiliki kemungkinan tinggi mengenai apa yang mungkin terjadi di masa yang akan datang. Dalam konteks data mining, prediksi melibatkan penerapan teknik statistik, machine learning, atau algoritma prediktif lainnya untuk mengidentifikasi pola atau tren dalam data yang dapat digunakan untuk memprediksi nilai atau kejadian di masa depan. Prediksi dapat diterapkan pada berbagai variabel atau parameter, seperti harga saham, kondisi cuaca, penjualan produk, dan bahkan perilaku pengguna. Ada beberapa algoritma yang dapat digunakan untuk melakukan proses prediksi diantaranya KNN, Naive Bayes, C4.5, Rough Set, SVM, dan lainnya [34], [35].

2.3 Agglomerative Hierarchical Clustering (AHC)

AHC merupakan sebuah algoritma atau metode dalam analisis kluster yang bertujuan untuk mengelompokkan data ke dalam kluster-kluster hierarkis. Pendekatan ini dimulai dengan memperlakukan setiap data sebagai sebuah kluster tunggal, dan kemudian secara bertahap menggabungkan kluster-kluster yang memiliki kesamaan tertinggi berdasarkan suatu metrik atau jarak yang ditentukan. Proses AHC dimulai dengan mengukur jarak antara setiap pasangan data, lalu menggabungkan dua kluster dengan jarak terdekat untuk membentuk satu kluster baru. Proses ini berlanjut dengan menggabungkan kluster-kluster yang semakin besar hingga semua data tergabung dalam satu kluster utama. Selama proses ini, terbentuklah sebuah struktur hirarkis yang menggambarkan hubungan antar kluster, biasanya dalam bentuk pohon dendrogram atau grafik [28], [36].

Salah satu kelebihan dari metode AHC adalah fleksibilitasnya dalam memilih metrik atau jarak yang digunakan untuk mengukur kesamaan antar data. Metrik tersebut dapat berupa jarak Euclidean, jarak Manhattan, atau jarak Mahalanobis, tergantung pada karakteristik data dan tujuan analisis. Selain itu, penggunaan AHC juga dapat dikontrol dengan memilih jumlah kluster yang diinginkan atau menggunakan kriteria pemotongan (cut-off) untuk menghentikan proses penggabungan kluster. Jarak antara dua cluster dapat dihitung menggunakan Euclidean Distance ataupun manhattan distance berikut rumusnya:

a. Menghitung matriks jarak

Euclidean Distance:

$$d(x_2, x_1) = \sqrt{\sum_{j=1}^d |x_j - y_j|^2} \quad (1)$$

Manhattan Distance:

$$D_{man}(x, y) = \sum_{j=1}^d |x_{2j} - y_{1j}| \quad (2)$$

b. Pengelompokan AHC

Proses pengclusteran AHC dapat dilakukan dengan 3 mode diantaranya single linkage yang digunakan dalam pengelompokan berdasarkan jarak terdekat (terkecil) antara satu cluster dengan cluster yang lain, complete linkage digunakan untuk mengelompokkan data cluster satu dengan yang lain berdasarkan jarak terjauh atau jarak maksimum antara elemen-elemen yang berbeda di dalam kluster tersebut. Pendekatan ini memusatkan perhatian pada mengukur jarak terbesar yang terjadi antara dua kluster, sehingga menghasilkan kluster yang memiliki perbedaan maksimum.



Mode ketiga yaitu average linkage yang digunakan dalam pengelompokkan berdasarkan nilai rata-rata dari dua cluster yang saling berpasangan [7][8][9][37]. Berikut persamaan yang dapat digunakan:

Single Linkage:

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \tag{3}$$

Complete Linkage:

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \tag{4}$$

Average Linkage:

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D \tag{5}$$

3. HASIL DAN PEMBAHASAN

Pengelompokkan provinsi yang ada di Indonesia akan dilakukan terhadap salah satu kriminal (kejahatan) yang terjadi yaitu pelecehan seksual. Pengelompokkan tersebut dilakukan agar masyarakat setempat ataupun pengunjung untuk lebih berwaspada (berhati-hati) dan meningkatkan keamanan diri agar terhindar dari kejahatan kriminal pelecehan seksual, adapun teknik pengelompokkan dalam data mining yang akan digunakan yaitu algoritma AHC. Metode ini dilakukan dengan membentuk cluster baru dari cluster yang telah ada (data awal dijadikan sebagai cluster-cluster), tahap awal dari

AHC yaitu menghitung matriks jarak menggunakan Euclidean distance kemudian dikelompokkan (membentuk cluster baru) menggunakan complete linkage. Pengelompokkan dilakukan terhadap 34 provinsi yang akan dijadikan sebagai sampel data dengan jumlah kasus pelecehan seksual terjadi selama 3 tahun yaitu tahun 2019, 2020 dan 2021. Berikut tabel 1 sampel data yang digunakan pada penelitian ini.

Tabel 1. Data Kriminal Pelecehan Seksual di Indonesia

Kode	Provinsi	Tahun 2019	Tahun 2020	Tahun 2021
P1	Aceh	0,99	1,58	2,54
P2	Sumatera Utara	0,43	0,15	1,77
P3	Sumatera Barat	0,59	0,48	1,54
P4	Riau	0,61	0	0,36
P5	Jambi	1,45	0,74	0,59
P6	Sumatera Selatan	0,57	1,11	0,73
P7	Bengkulu	1,87	0	0,56
P8	Lampung	0,54	0,23	0,49
P9	Kep.Bangka Belitung	2,23	0	4,48
P10	Kepulauan Riau	0	0,02	0,67
P11	DKI Jakarta	1,64	2,68	0
P12	Jawa Barat	0,91	1,88	3,7
P13	Jawa Tengah	0,81	0,74	0,49
P14	D.I Yogyakarta	3,1	0,4	1,71
P15	Jawa Timur	3,84	1,2	1,69
P16	Banten	1,08	3,91	2,76
P17	Bali	2,65	0	0
P18	Nusa Tenggara Barat	0,57	0,74	0,27
P19	Nusa Tenggara Timur	1,82	1,4	1,99
P20	Kalimantan Barat	0,35	0,29	0,72
P21	Kalimantan Tengah	3,47	1,88	0,93
P22	Kalimantan Selatan	1,03	0	2,98
P23	Kalimantan Timur	0,7	0,45	0,35
P24	Kalimantan Utara	0	2,86	0
P25	Sulawesi Utara	0,84	0,29	0,48
P26	Sulawesi Tengah	0,61	1,3	0,27
P27	Sulawesi Selatan	0,52	0,48	2,28
P28	Sulawesi Tenggara	0,27	2,01	1,41
P29	Gorontalo	0,66	4,59	0
P30	Sulawesi Barat	0	1,54	0
P31	Maluku	0,04	2,95	1,76
P32	Maluku Utara	1,44	0	3,16
P33	Papua Barat	1,22	2,44	1,6
P34	Papua	0,45	4,44	4,11



3.1 Penerapan Metode AHC

a. Menghitung jarak matriks dengan menggunakan rumus euclidean berikut

$$d(x_2, x_1) = \sqrt{\sum_{j=1}^d |x_j - y_j|^2}$$

$$d(A1, A2) = \sqrt{|0,99 - 0,43|^2 + |1,58 - 0,15|^2 + |2,54 - 1,77|^2}$$

$$d(A1, A3) = \sqrt{|0,99 - 0,59|^2 + |1,58 - 0,48|^2 + |2,54 - 1,54|^2}$$

$$d(A1, A4) = \sqrt{|0,99 - 0,61|^2 + |1,58 - 0|^2 + |2,54 - 0,36|^2}$$

$$d(A1, A5) = \sqrt{|0,99 - 1,45|^2 + |1,58 - 0,74|^2 + |2,54 - 0,59|^2}$$

$$d(A1, A6) = \sqrt{|0,99 - 0,57|^2 + |1,58 - 1,11|^2 + |2,54 - 0,73|^2}$$

$$d(A1, A7) = \sqrt{|0,99 - 1,87|^2 + |1,58 - 0|^2 + |2,54 - 0,56|^2}$$

Lakukan perhitungan pencarian jarak matriks yang disebutkan di atas dengan menggunakan rumus euclidean distance dari kode A1, A34 hingga kode A33, A34. Untuk hasil perhitungan yang telah dilakukan dapat dilihat pada tabel 2 berikut.

Tabel 2. Matriks Jarak Euclidean

Kode	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A33	A34
A1	0											
A2	1,7	0										
A3	1,5	0,4	0									
A4	2,7	1,4	1,3	0								
A5	2,2	1,7	1,3	1,1	0							
A6	1,9	1,4	1	1,2	1	0						
A7	2,7	1,9	1,7	1,3	0,9	1,7	0					
A8	2,5	1,3	1,1	0,3	1	0,9	1,4	0				
A9	2,8	3,3	3,4	4,4	4	4,2	3,9	4,3	0			
A10	2,6	1,2	1,1	0,7	1,6	1,2	1,9	0,6	4,4	0		
A11	2,8	3,3	2,9	2,9	2	2	2,7	2,7	5,3	3,2		
A12	1,2	2,6	2,6	3,8	3,4	3,1	3,8	3,6	2,4	3,7		
A13	2,2	1,5	1,1	0,8	0,6	0,5	1,3	0,6	4,3	1,1		
A14	2,6	2,7	2,5	2,9	2	2,8	1,7	2,8	2,9	3,3		
A15	3,0	3,6	3,3	3,7	2,7	3,4	2,6	3,6	3,4	4,1		
A16	2,3	3,9	3,7	4,6	3,9	3,5	4,6	4,4	4,4	4,5		
A17	3,4	2,8	2,6	2,1	1,5	2,5	1	2,2	4,5	2,7		
A18	2,5	1,6	1,3	0,7	0,9	0,6	1,5	0,6	4,6	1		
A19	1,0	1,9	1,6	2,5	1,6	1,8	2	2,3	2,9	2,6		
A20	2,3	1,1	0,9	0,5	1,2	0,8	1,6	0,3	4,2	0,4		
...		
...		
A33	1,3	2,4	2,1	2,8	2	1,7	2,7	2,6	3,9	2,9	0	
A34	3,3	4,9	4,7	5,8	5,2	4,7	5,9	5,6	4,8	5,6	3,3	0

Berdasarkan tabel 2, pilih jarak euclidean terkecil yang dijadikan sebagai cluster awal, nilai matriks terkecil (minimum) ada pada matriks A23 A25 dengan nilai 0,2, cluster yang berada didalam cluster A23A25 dinyatakan sebagai cluster yang tingkat kriminalnya sedikit atau dinyatakan sebagai provinsi yang lebih aman.

b. Pada cluster yang telah ditentukan sebagai satu cluster maka digabung kodenya menjadi A23A25, lakukan pengelompokkan selanjutnya dengan menggunakan mode *complete linkage* dimana jarak antara cluster satu dengan cluster selanjutnya ditentukan berdasarkan jarak terjauh. Berikut proses pencarian nilai maksimal antara satu cluster dengan cluster berikutnya (antara dua cluster) yang akan digabungkan.

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D$$

$$d_{A23A25,A1} = \max\{D(A23, A1); D(A25, A1)\} = \text{MAX}\{2,5; 2,4\} = 2,5$$

$$d_{A23A25,A2} = \max\{D(A23, A2); D(A25, A2)\} = \text{MAX}\{1,5; 1,4\} = 1,5$$

$$d_{A23A25,A3} = \max\{D(A23, A3); D(A25, A3)\} = \text{MAX}\{1,2; 1,1\} = 1,2$$

$$d_{A23A25,A4} = \max\{D(A23, A4); D(A25, A4)\} = \text{MAX}\{0,5; 0,4\} = 0,5$$

$$d_{A23A25,A5} = \max\{D(A23, A5); D(A25, A5)\} = \text{MAX}\{0,8; 0,8\} = 0,8$$

$$d_{A23A25,A6} = \max\{D(A23, A6); D(A25, A6)\} = \text{MAX}\{0,8; 0,9\} = 0,9$$

$$d_{A23A25,A7} = \max\{D(A23, A7); D(A25, A7)\} = \text{MAX}\{1,3; 1,1\} = 1,3$$

Lakukan perhitungan compkate linkage diatas terhadap cluster lainnya hingga selesai, carilah nilai terkecil dari cluster yang terhubung dengan A23 A25. Tidak perlu mencari nilai terbesar untuk cluster lainnya (selain cluster yang terhubung dengan A23 A25), hanya perlu Memindahkan Matriks Jarak Euclidean pada tabel 1. Anda dapat melihat hasil penyelesaian pada tabel 3 matriks pembaharuan yang disediakan berikut.

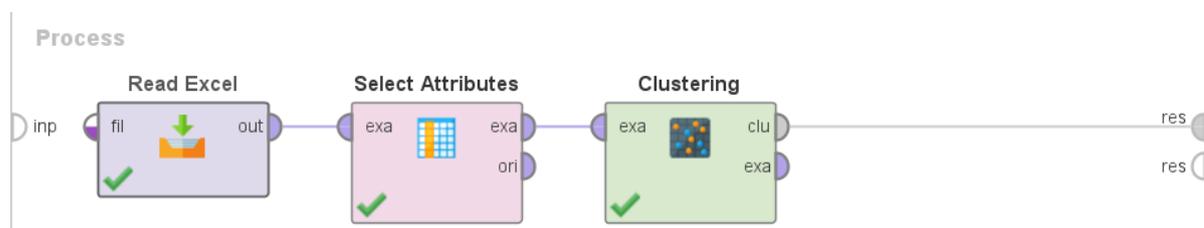
Tabel 3. Matriks Pembentukan cluster pertama

Kode	A23,A25	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A33	A34
A23,A25	0,0												
A1	2,5	0											
A2	1,5	0,4	0										
A3	1,2	1,4	1,3	0									
A4	0,5	1,7	1,3	1,1	0								
A5	0,8	1,4	1	1,2	1	0							
A6	0,9	1,9	1,7	1,3	0,9	1,7	0						
A7	1,3	1,3	1,1	0,3	1	0,9	1,4	0					
A8	0,3	3,3	3,4	4,4	4	4,2	3,9	4,3	0				
A9	4,4	1,2	1,1	0,7	1,6	1,2	1,9	0,6	4,4	0			
A10	0,9	3,3	2,9	2,9	2	2	2,7	2,7	5,3	3,2	0		
A11	2,6	2,6	2,6	3,8	3,4	3,1	3,8	3,6	2,4	3,7	3,9		
A12	3,6	1,5	1,1	0,8	0,6	0,5	1,3	0,6	4,3	1,1	2,2		
A13	0,5	2,7	2,5	2,9	2	2,8	1,7	2,8	2,9	3,3	3,2		
A14	2,8	3,6	3,3	3,7	2,7	3,4	2,6	3,6	3,4	4,1	3,1		
A15	3,5	3,9	3,7	4,6	3,9	3,5	4,6	4,4	4,4	4,5	3,1		
A16	4,3	2,8	2,6	2,1	1,5	2,5	1	2,2	4,5	2,7	2,9		
A17	2,0	1,6	1,3	0,7	0,9	0,6	1,5	0,6	4,6	1	2,2		
A18	0,6	1,9	1,6	2,5	1,6	1,8	2	2,3	2,9	2,6	2,4		
A19	2,2	1,1	0,9	0,5	1,2	0,8	1,6	0,3	4,2	0,4	2,8		
A20	0,5	3,6	3,3	3,5	2,3	3	2,5	3,4	4,2	3,9	2,2		
...		
...		
A33	2,5	1,3	2,4	2,1	2,8	2,0	1,7	2,7	2,6	3,9	2,9	0	
A34	5,5	3,3	4,9	4,7	5,8	5,2	4,7	5,9	5,6	4,8	5,6	3,3	0

Melalui analisis tabel 3, telah berhasil membentuk sebuah cluster baru dengan menghitung jarak terjauh antara cluster-cluster yang ada (maksimal). Cluster baru yang terbentuk adalah A23 A25. Selanjutnya, dalam proses pencarian cluster berikutnya, lakukan pencarian nilai minimum dari matriks tabel 3. Untuk mencari nilai cluster yang terhubung dengan cluster baru tersebut maka lakukan sesuai langkah 3 dan kembali ke langkah 2. Langkah ini memungkinkan kami mengelompokkan kode provinsi berdasarkan cluster yang diinginkan. Untuk melihat hasil akhir dengan lebih jelas, Anda dapat menggunakan perangkat lunak RapidMiner.

3.2 Pengujian Metode AHC

Proses pengujian metode AHC dilakukan dengan menggunakan perangkat lunak RapidMiner. Tujuan pengujian ini adalah untuk mendapatkan hasil akhir (pengelompokkan) secara efisien dan menghindari kesalahan serta kesalahan manusia yang mungkin terjadi dalam proses pengelompokkan manual. Berikut adalah langkah-langkah pembentukan cluster menggunakan RapidMiner.



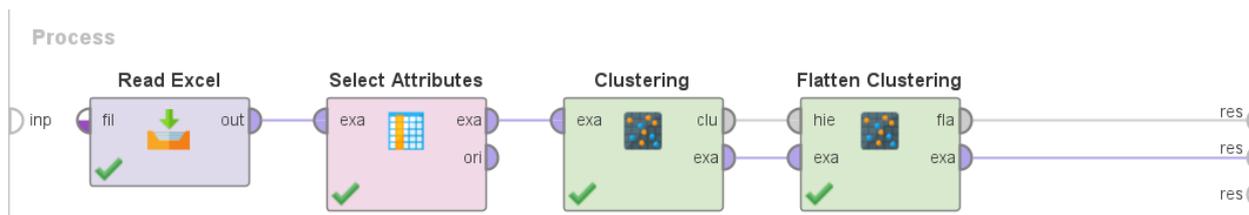
Gambar 2. Proses dataset inputan file

Ilustrasi dalam Gambar 2 menjelaskan langkah-langkah penggunaan RapidMiner. Tahap awal yaitu memasukkan file data yang akan diproses sebagai input (file excel). Kemudian, digunakan alat pemilihan atribut (select atribut) untuk memilih atribut yang akan digunakan dari file tersebut. Penggunaan alat pemilihan atribut tidak terlalu penting jika kriteria data awal sudah ditentukan sebelumnya. Namun, jika file tersebut mengandung banyak informasi yang tidak relevan, alat pemilihan atribut dapat mempermudah proses tersebut. Untuk pengklasteran, digunakan alat pengelompokan (clustering tool) dengan metode agglomerative dan metode penghitungan jarak maksimal, yang disebut metode complete linked. Metode ini melibatkan perhitungan jarak hierarkis antara provinsi satu dengan provinsi lainnya. Setelah menjalankan proses RapidMiner, akan terbentuk dendrogram yang menggambarkan hasil klusterisasi.



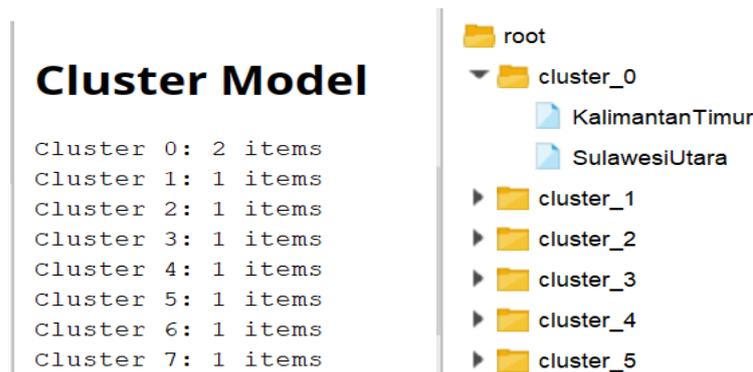
Gambar 3. Dendrogram AHC

Dalam gambar 3, terlihat bahwa dendrogram AHC menunjukkan pembentukan hierarki berdasarkan jarak terjauh antara provinsi yang menghasilkan pembentukan cluster-cluster baru. Untuk mempermudah penyelesaian metode AHC, penelitian ini akan membentuk 3 cluster. Oleh karena itu, perlu menambahkan alat penggabungan kluster (flatten clustering tool) pada input untuk menentukan jumlah cluster yang akan dibentuk. Berikut adalah proses input setelah menambahkan alat flatten clustering.



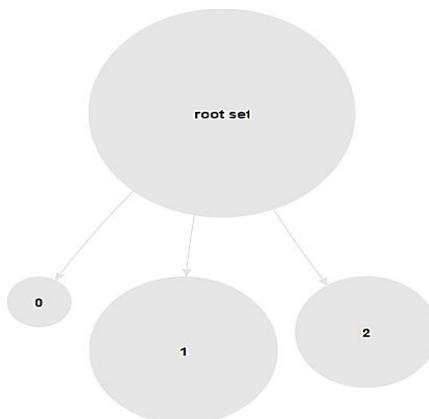
Gambar 4. Proses dataset inputan file

Ilustrasi dalam Gambar 4 menunjukkan proses input lengkap yang akan dilakukan, mulai dari input data dari file Excel, pemilihan atribut yang digunakan, penerapan metode clustering (agglomerative hierarchical clustering) dengan memilih mode pengelompokan complete linked, serta penggunaan flatten clustering untuk menentukan jumlah cluster yang akan terbentuk, sebagai pembuktian kesesuaian perhitungan manual yang telah dilakukan sebelumnya dengan penerapan rapidminer, maka peneliti menguji aplikasi rapidminer dengan membentuk 33 cluster, dimana satu cluster (cluster) pertama yang dibentuk harus sesuai (sama) antara perhitungan manual dengan penggunaan aplikasi rapid miner. Berikut tampilan cluster model yang terbentuk sebanyak 34 cluster.



Gambar 5. Pembentukan cluster pertama

Berdasarkan gambar 5, terlihat bahwa cluster pertama (cluster 0) menampung provinsi kalimantan timur dan sulawesi utara. Jika dilihat pada perhitungan manual, diperoleh A23 dan A25 sebagai cluster pertama yang terbentuk yaitu provinsi kalimantan timur dan sulawesi utara. Sehingga dapat diputuskan bahwa proses manual dan aplikasi rapid miner sesuai dan dapat dilanjutkan menggunakan rapidminer. Sekarang berfokus pada pembentukan 3 cluster dimana pada saat di jalan aplikasi rapidminer maka terbentuk kumpulan pohon akar (rootset trees) yang menggambarkan hasil dari proses tersebut.



Gambar 6. Proses dataset inputan file

Dari ilustrasi dalam Gambar 6, terlihat bahwa cluster 0 memiliki jumlah provinsi yang paling sedikit, sementara jumlah provinsi terbanyak ada di cluster 1. Informasi mengenai jumlah provinsi dalam setiap cluster dapat ditemukan pada gambar terlampir.

Cluster Model

Cluster 0: 5 items
 Cluster 1: 17 items
 Cluster 2: 12 items
 Total number of items: 34

Gambar 7. Data cluster model

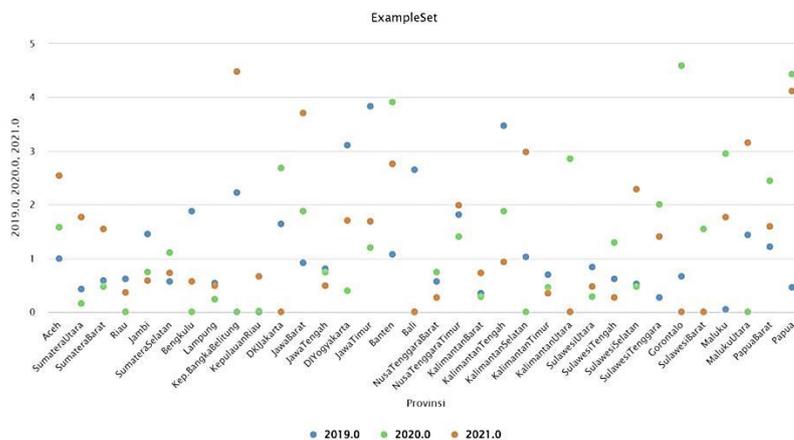
Melihat dari ilustrasi dalam Gambar 7, terlihat bahwa terdapat 5 provinsi yang dikelompokkan ke dalam cluster 0, kemudian terdapat 17 provinsi dalam cluster 1, dan 12 provinsi dalam cluster 2. Untuk informasi detail mengenai pengelompokkan tersebut, dapat dilihat pada tabel berikut.

Tabel 4. Cluster terbentuk setiap provinsi

No	Provinsi	2019	2020	2021	Cluster
1	Aceh	0,99	1,58	2,54	1
2	SumateraUtara	0,43	0,15	1,77	2
3	SumateraBarat	0,59	0,48	1,54	2
4	Riau	0,61	0	0,36	1
5	Jambi	1,45	0,74	0,59	2
6	SumateraSelatan	0,57	1,11	0,73	1
7	Bengkulu	1,87	0	0,56	2
8	Lampung	0,54	0,23	0,49	1
9	Kep.BangkaBelitung	2,23	0	4,48	2
10	KepulauanRiau	0	0,02	0,67	1
11	DKIJakarta	1,64	2,68	0	0
12	JawaBarat	0,91	1,88	3,7	1
13	JawaTengah	0,81	0,74	0,49	1
14	DIYogyakarta	3,1	0,4	1,71	2
15	JawaTimur	3,84	1,2	1,69	2
16	Banten	1,08	3,91	2,76	0
17	Bali	2,65	0	0	2
18	NusaTenggaraBarat	0,57	0,74	0,27	1
19	NusaTenggaraTimur	1,82	1,4	1,99	1

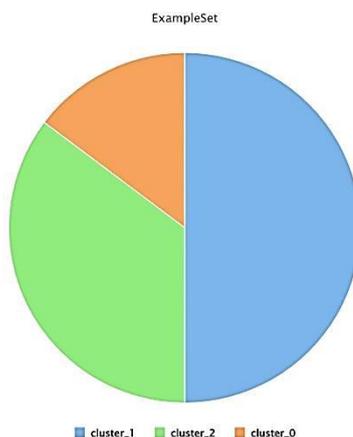
20	KalimantanBarat	0,35	0,29	0,72	1
21	KalimantanTengah	3,47	1,88	0,93	2
22	KalimantanSelatan	1,03	0	2,98	2
23	KalimantanTimur	0,7	0,45	0,35	1
24	KalimantanUtara	0	2,86	0	0
25	SulawesiUtara	0,84	0,29	0,48	1
26	SulawesiTengah	0,61	1,3	0,27	1
27	SulawesiSelatan	0,52	0,48	2,28	2
28	SulawesiTenggara	0,27	2,01	1,41	1
29	Gorontalo	0,66	4,59	0	0
30	SulawesiBarat	0	1,54	0	1
31	Maluku	0,04	2,95	1,76	1
32	MalukuUtara	1,44	0	3,16	2
33	PapuaBarat	1,22	2,44	1,6	1
34	Papua	0,45	4,44	4,11	0

Berdasarkan tabel 4, dapat dijelaskan bahwa provinsi yang dikelompokkan kedalam cluster pertama (0) sebanyak 5 provinsi yang terdiri dari DKI Jakarta, Banten, Kalimantan Utara, Gorontalo dan Papua dan juga dinyatakan sebagai provinsi rawan terjadi pelecehan seksual. Cluster 1 terdiri dari 17 provinsi diantaranya Aceh, Riau, Sumatera Selatan, Lampung, Kepulauan Riau, Jawa Barat, Jawa Tengah, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat, Maluku dan Papua Barat. Cluster 2 terdiri dari 12 provinsi diantaranya Sumatera Utara, Sumatera Barat, Jambi, Bengkulu, Kep. Bangka Belitung, DI Yogyakarta, Jawa Timur, Bali, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat dan Papua.



Gambar 8. Scatter clustering complete link

Berdasarkan gambar 8 dapat dijelaskan bahwa provinsi yang terjadi kriminal pelecehan seksual pada tahun 2019 ditandai dengan warna biru, untuk kejadian kriminal di tahun 2020 ditandai dengan warna hijau dan kriminal pelecehan seksual terjadi di tahun 2021 ditandai dengan warna orange.



Gambar 9. Visualisasi Cluster

Berdasarkan gambar 9, dapat dijelaskan bahwa cluster dengan jumlah provinsi terbanyak ada pada cluster 1 yaitu cluster yang ditandai berwarna biru dengan jumlah provinsi sebanyak 17 provinsi dan persentase sebesar 50% dari keseluruhan provinsi (provinsi yang ada di Indonesia). Cluster 2 berwarna hijau dengan jumlah provinsi sebanyak 12 provinsi dan persentase sebesar 35,29%. Dan cluster 0 dengan jumlah provinsi sedikit yaitu 5 dan persentasenya hanya 14,71%

4. KESIMPULAN

Kesimpulan dari penelitian ini yaitu diharapkan pelecehan seksual yang sering terjadi terkhusus terhadap wanita agar dan juga pembaca agar dapat dilakukan pencegahan terhadap kejahatan pelecehan seksual dengan tindakan yang cepat di tempat-tempat yang sering terjadi pelecehan seksual melalui sosialisasi, peningkatan keamanan, dan penyediaan tempat konsultasi seperti psikolog. Metode data mining, dengan menggunakan metode pengelompokan AHC dan mode complete linkage, digunakan untuk mengidentifikasi tempat-tempat yang rawan terjadinya pelecehan seksual berdasarkan data-data sebelumnya. Proses penelitian ini melibatkan kombinasi antara analisis manual dan penggunaan aplikasi RapidMiner untuk membentuk kluster baru. Hasil penelitian menunjukkan adanya tiga kluster, dengan 5 provinsi dalam kluster 0, 17 provinsi dalam kluster 1, dan 12 provinsi dalam kluster 2.

REFERENCES

- [1] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," *JUITA J. Inform.*, vol. 7, no. 2, pp. 71–79, 2019.
- [2] M. Zunaidi, A. H. Nasyuha, and S. M. Sinaga, "Penerapan Data Mining Untuk Memprediksi Pertumbuhan Jumlah Penderita Human Immunodeficiency Virus (HIV) Menggunakan Metode Multiple Linier Regression (Studi Kasus Dinas Kesehatan Provinsi Sumatera Utara)," *J. Teknol. Sist. Inf. Dan Sist. Komput. TGD*, vol. 3, no. 1, pp. 137–147, 2020.
- [3] H. F. Sinurat, S. D. Nasution, and A. Fau, "Penerapan Metode Self Organizing Maps (SOM) Dalam Rekomendasi Jurusan Calon Mahasiswa Baru (Studi Kasus: Universitas Imelda Medan)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, 2020.
- [4] 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.
- [5] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021.
- [6] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021.
- [7] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020.
- [8] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–9.
- [9] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13657–13665.
- [10] R. T. Aldisa, "Data Mining Penentuan Jurusan Siswa Menggunakan Metode Agglomerative Hierarchical Clustering (AHC)," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 2, pp. 873–880, 2023.
- [11] R. Rahma and R. Mufidah, "Pengelompokan daerah rawan kekerasan terhadap perempuan dan anak di Jawa Barat menggunakan algoritma k-means," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 3, pp. 850–857, 2022.
- [12] L. Zahrotun, S. H. Nugroho, U. Linarti, and A. H. S. Jones, "Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linkage," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 2, pp. 255–265, 2023.
- [13] S. M. Dewi, A. P. Windarto, I. S. Damanik, and H. Satria, "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah," *Semin. Nas. Sains Teknol. Inf.*, pp. 620–625, 2019.
- [14] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021.
- [15] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.
- [16] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth);," *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.
- [17] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, pp. 133–137, 2021.
- [18] S. S. S. A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020.
- [19] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," *Bull. Inf. Technol.*, vol. 1, no. 2, pp. 83–91, 2020.
- [20] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018.
- [21] A. Wanto et al., *Data Mining: Algoritma dan Implementasi*. Yayasan kita menulis, 2020.
- [22] N. L. W. S. R. Ginantra et al., *Data mining dan penerapan algoritma*. Yayasan Kita Menulis, 2021.
- [23] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021.



- [24] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 219–225, 2021.
- [25] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–432, 2019.
- [26] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, pp. 133–144, 2021.
- [27] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusantara*, vol. 2, no. 4, pp. 191–197, 2021.
- [28] B. Harli Trimulya Suandi As and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud)," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
- [29] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori Dan Fpgrowth Sebagai Dasar Pertimbangan Penentuan Paket Sembako," 2020.
- [30] K. Erwansyah, B. Andika, and R. Gunawan, "Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritma Apriori Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 4, no. 1, pp. 148–161, 2021.
- [31] A. S. L. T. H. Hafizah, "Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022.
- [32] Y. L. Nainel, E. Buulolo, and I. Lubis, "Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk)," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 214, 2020.
- [33] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, pp. 55–61, 2019.
- [34] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [35] S. U. Putri, E. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4. 5," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
- [36] R. A. Setyawan and R. M. Fadilla, "Klasterisasi media pembelajaran daring di era pandemi COVID-19 menggunakan metode Agglomerative," *Inf. Interaktif*, vol. 5, no. 3, 2020.
- [37] N. K. Zuhail, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022.