ANALISA PEMERATAAN IMUNISASI CAMPAK PADA ANAK SEKOLAH DI JAKARTA DENGAN ALGORITMA CLUSTEING HIERARKI DAN KLASIFIKASI STANDAR

Freda Widya Artanti, Nur Atika, Kharisma Putri Sholekha,
Zahra Shabrina Aderi, Aisyah Muti Yanuariska
Program Studi Sistem Informasi, Universitas Amikom Purwokerto
Jl. Letjend Pol. Soemarto No.127, Watumas, Purwanegara, Kec. Purwokerto Utara
Kabupaten Banyumas, Jawa Tengah 53127
fredawidya09@gmail.com

ABSTRAK

Campak adalah sebuah penyakit menular yang dapat menyerang anak-anak yang terjadi akibat virus. Penularan campak terjadi melalui perantaraan batuk dan bersin dari manusia ke manusia. Campak dapat menyebabkan kecacatan dan bahkan kematian, sehingga perlu dilakukan upaya lebih lanjut untuk mengatasi masalah ini. Salah satu langkah dalam mengendalikan penyebaran campak adalah melalui program imunisasi yang bertujuan untuk mencegah perkembangan dari penyakit tersebut. Dalam penelitian ini, data yang akan digunakan berasal dari data.jakarta.go.id. Penelitian ini memanfaatkan teknik data mining dengan menggunakan algoritma clustering hierarki. Metode clustering hierarki digunakan untuk mengelompokkan data menjadi beberapa kelompok. Keunggulan metode ini termasuk kemampuannya dalam mengatasi masalah yang sering dihadapi oleh metode K-Means, yaitu sensitivitas terhadap data yang berbeda. Algoritma ini juga memiliki keunggulan lain, yaitu hasil clustering tidak tergantung pada urutan entri dalam dataset. Dengan menerapkan metode clustering hierarki, penelitian ini bertujuan untuk mengidentifikasi data persentase imunisasi campak berdasarkan provinsi, sehingga provinsi dapat dikelompokkan berdasarkan tingkat imunisasi mereka. Dari penelitian ini didapatkan, dari satu set pelatihan (80% dari catatan) dan satu set validasi (20% dari catatan) diklasifikasikan dengan benar setelah pelatihan sementara akurasi klasifikasi dataset pelatihan berkinerja sangat baik (sekitar 84% untuk random forest) akurasi untuk dataset validasi turun dibawah 50%. Bahwa analisis dalam ruang atribut (fitur) lengkap mungkin tidak efektif, sehingga dilakukan eksplorasi analisis klaster hierarki untuk mempelajari mengelompokan data hasil pengobatan imunisasi pada anak sekolah di daerah ibukota jakarta. Hasilnya bahwa segmentasi data untuk hasil pengobatan diikuti oleh pengelompokan subruang memungkinkan untuk mengidentifikasi kelompok data yang relevan, yang dapat membentuk dasar untuk generalisasi pengetahuan medis. Harapannya, penelitian ini dapat memberikan informasi yang berharga kepada pemerintah mengenai pengelompokan data imunisasi campak di antara anak-anak sekolah di Indonesia, khususnya di ibukota Jakarta.

***Kata kunci**: Clustering Hierarki, Data Mining, Clustering, Immunization*

1. PENDAHULUAN

Campak adalah penyakit menular yang bisa menyerang anak-anak maupun orang dewasa dan disebabkan oleh virus. Penyakit ini dapat menular melalui kontak langsung seperti batuk dan bersin. Berdasarkan data dari Organisasi Kesehatan Dunia [1], Indonesia masuk dalam 10 negara dengan jumlah kasus campak tertinggi di dunia. Kementerian Kesehatan mencatat bahwa jumlah kasus campak di Indonesia sangat besar dan mengalami peningkatan dalam lima tahun terakhir. Beberapa komplikasi yang dapat terjadi pada penderita campak termasuk radang paru-paru, infeksi telinga, diare, dan radang otak. Penyakit campak merupakan penyebab utama kematian anak akibat penyakit yang bisa dicegah melalui vaksinasi.

Menurut informasi yang diberikan oleh Kementerian Kesehatan Republik Indonesia pada tahun 2016 [2] "Cakupan imunisasi campak di Indonesia mencapai 84% dan diklasifikasikan sebagai sedang jika dibandingkan dengan negara-negara lain di kawasan Asia Tenggara." Pada tahun 2020, Indonesia turut serta dalam program eliminasi campak yang menetapkan target cakupan minimal 95% di setiap wilayah secara merata sebagai prioritas utama pemerintah [3]. Program imunisasi campak merupakan salah satu langkah dalam upaya melawan penyakit campak, di mana virus penyakit yang telah dilemahkan dimasukkan ke dalam tubuh melalui suntikan atau pemberian melalui mulut [4]. Tujuan dari program ini adalah untuk meningkatkan kekebalan tubuh terhadap penyakit campak. Namun, disayangkan bahwa di Indonesia, tingkat kasus campak pada anak-anak usia dini masih relatif tinggi. Diperkirakan sekitar 30.000 anak di Indonesia kehilangan nyawa setiap tahunnya akibat komplikasi yang disebabkan oleh campak. Data ini juga diperkuat oleh fakta bahwa lebih dari 95% kematian akibat campak terjadi di berbagai negara, termasuk Amerika Serikat, Inggris, Yaman, Amerika Latin, Myanmar, Afrika, Chad, Afrika Tengah, Liberia, Guinea, dan Indonesia. Oleh karena itu, pemerintah

memiliki tanggung jawab dan komitmen untuk menghapus penyakit campak [5]. Untuk mencapai tujuan ini, terutama dalam memutus mata rantai penularan campak pada anak-anak usia dini, diperlukan tingkat imunisasi minimal sebesar 95%. Untuk mencapai target ini, pemerintah perlu memiliki data yang akurat untuk melakukan pemetaan tingkat imunisasi campak [6]. Data ini diperoleh dari Badan Pusat Statistik (BPS) mengenai persentase imunisasi berdasarkan wilayah di Jakarta pada tahun 2019.

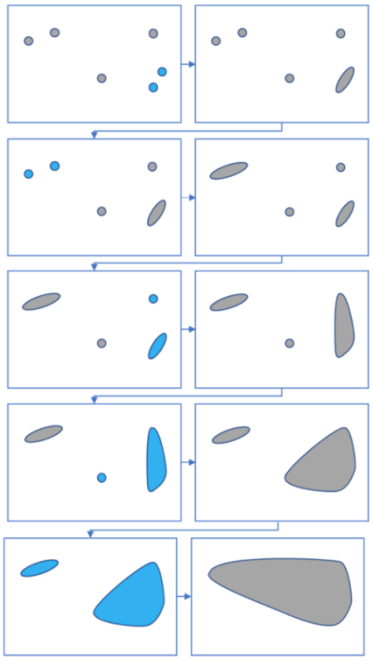
Metode clustering hierarki bisa digunakan pada data persentase imunisasi campak berdasarkan wilayah, sehingga memungkinkan kita untuk menentukan pengelompokan wilayah berdasarkan data ini [7]. Dengan melihat data pengelompokan tersebut, kita dapat mengidentifikasi karakteristik masing-masing cluster, termasuk cluster dengan tingkat imunisasi campak rendah, sedang, dan tinggi di setiap wilayah. Harapannya, penelitian ini dapat memberikan informasi yang berharga kepada pemerintah tentang pengelompokan data imunisasi campak di Indonesia, yang akan berkontribusi pada upaya pemerintah dalam mendistribusikan imunisasi campak secara merata di Ibukota Jakarta.

2. TINJAUAN PUSTAKA

2.1. Clustering Hierarki

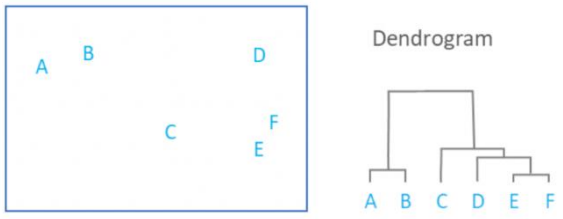
Clustering hierarkis, atau sering disebut analisis pengelompokan hierarkis, merupakan suatu metode algoritma yang mengelompokkan objek yang memiliki kesamaan karakteristik ke dalam kelompok yang disebut sebagai "cluster". Hasil akhir dari proses ini adalah kumpulan cluster, di mana setiap cluster memiliki perbedaan yang signifikan dengan cluster lainnya, dan objek-objek dalam setiap cluster memiliki kemiripan karakteristik yang mencolok satu sama lain [8].

Hierarchical clustering dapat dilakukan dengan matriks jarak atau data mentah. Ketika data mentah disediakan, perangkat lunak akan secara otomatis menghitung matriks jarak di latar belakang. Matriks jarak di bawah ini menunjukkan jarak antara enam objek [9]. Hierarchical clustering dimulai dengan memperlakukan setiap pengamatan sebagai klaster yang terpisah. Kemudian berulang kali menjalankan dua langkah berikut: (1) mengidentifikasi dua cluster yang paling dekat, dan (2) menggabungkan dua cluster yang paling mirip. Proses iteratif ini terus berlanjut sampai semua cluster digabungkan menjadi satu [10]. Hal ini diilustrasikan dalam diagram di bawah ini.



Gambar 1. Cara kerja hierarchical clustering

Output utama dari Hierarchical Clustering adalah dendrogram, yang menunjukkan hubungan hierarkis antara cluster:



Gambar 2. Hasil umum dari hierarchical clustering

Pada gambar contoh di atas, jarak antara kedua cluster yang telah terhitung berdasarkan panjang garis lurus yang ditarik dari satu cluster ke cluster lainnya. Ini biasanya disebut sebagai jarak Euclidean [11]. Banyak metrik jarak lainnya telah dikembangkan.

Pilihan metrik jarak harus dibuat berdasarkan pertimbangan teoritis dari domain studi. Artinya, metrik jarak perlu mendefinisikan kesamaan dengan cara yang masuk akal untuk bidang studi [12]. Misalnya, jika mengelompokkan situs kejahatan di kota, jarak blok kota mungkin tepat. Atau, lebih baik lagi, waktu yang dibutuhkan untuk melakukan perjalanan antar setiap lokasi. Di mana tidak ada pembenaran teoretis untuk alternatif, Euclidean umumnya harus lebih disukai, karena biasanya merupakan ukuran jarak yang tepat di dunia fisik [13].

Setelah menetapkan metrik jarak, diperlukan penentuan sumber yang digunakan untuk mengukur jarak tersebut. Sebagai contoh, jarak dapat diukur antara dua elemen yang memiliki kesamaan tertinggi dalam sebuah cluster (single-linkage), dua elemen yang paling tidak mirip dalam sebuah cluster

(complete-linkage), pusat cluster (mean atau average-linkage), atau berdasarkan kriteria-kriteria lainnya. Ada banyak kriteria keterkaitan yang telah dikembangkan.

Sama seperti ketika memilih metrik jarak, penentuan kriteria keterkaitan juga harus dilakukan berdasarkan pertimbangan teoritis yang relevan dengan domain aplikasi. Isu teoritis utama adalah faktor-faktor apa yang berkontribusi terhadap variasi yang diamati. Misalnya, dalam arkeologi, kami mengharapkan variasi terjadi melalui inovasi dan sumber daya alam, jadi mencari tahu apakah dua kelompok artefak serupa mungkin masuk akal berdasarkan mengidentifikasi anggota cluster yang paling mirip [14].

Di mana tidak ada pembenaran teoretis yang jelas untuk pemilihan kriteria keterkaitan, metode Ward adalah standar yang masuk akal. Metode ini menentukan pengelompokan pengamatan berdasarkan perhitungan reduksi jumlah kuadrat jarak antara tiap pengamatan dan nilai rata-rata pengamatan dalam suatu cluster. Pendekatan ini sering kali akurat karena konsep jarak ini sejalan dengan asumsi standar dalam pengukuran perbedaan antar kelompok dalam analisis statistik, seperti ANOVA atau MANOVA.

2.2. Equity

Dalam konteks kesehatan atau kesejahteraan sosial, equitas dapat merujuk pada upaya untuk mencapai distribusi yang adil dan setara dari layanan kesehatan, sumber daya, dan manfaat kesehatan di antara berbagai kelompok dalam masyarakat. Hal ini seringkali berkaitan dengan mengatasi disparitas kesehatan dan memastikan bahwa semua individu memiliki akses yang setara terhadap layanan kesehatan. Secara umum, "equitas imunisasi" mungkin merujuk pada usaha untuk mencapai distribusi yang adil dan setara dari manfaat vaksinasi atau kekebalan imun di antara populasi. Dalam konteks kesehatan masyarakat, equitas imunisasi dapat mencakup beberapa aspek, seperti, Akses yang Merata, Distribusi yang Adil, Pemerataan Kecukupan Imunisasi, Pertimbangan Etika dan Sosial. Dapat disimpulkan bahwa, equitas mencerminkan prinsip-prinsip keadilan, kesetaraan, dan perlakuan yang adil dalam berbagai aspek kehidupan. Implementasi equitas dapat melibatkan upaya aktif untuk mengatasi ketidaksetaraan dan memastikan bahwa hak-hak dan peluang didistribusikan secara merata di antara semua anggota masyarakat.

2.3. Standar Classification

Standar klasifikasi mengacu pada seperangkat kriteria, aturan, atau sistem yang digunakan untuk mengkategorikan dan mengelompokkan entitas atau fenomena tertentu. Tujuan utama dari standar klasifikasi adalah menciptakan kerangka kerja yang seragam dan konsisten untuk pengelompokan yang diterima secara luas. Standar klasifikasi melibatkan klasifikasi barang dan layanan, industri, atau bahkan data geografis. seperangkat pedoman atau aturan yang digunakan untuk mengorganisir dan mengelompokkan entitas atau fenomena tertentu ke dalam kategori atau kelas yang sesuai. Ini membantu menciptakan struktur yang jelas dan konsisten, memudahkan pemahaman, analisis, dan pertukaran informasi.

2.4. Penelitian Terdahulu

Berdasarkan uraian di atas, dalam ilmu komputer, terdapat berbagai cabang kecerdasan buatan yang dapat mengatasi permasalahan yang bersifat kompleks. Di antara cabang-cabang tersebut adalah sistem pendukung keputusan, sistem pakar, serta teknik data mining, yang merupakan beberapa contoh yang dapat digunakan dalam berbagai penelitian. Misalnya, penelitian yang mengkaji data mining telah memberikan evaluasi berdasarkan indeks yang mencakup tingkat ketersediaan sarana kesehatan di Desa/Kelurahan di seluruh provinsi Indonesia. Hasil penelitian tersebut menunjukkan bahwa empat provinsi, yakni Sumatera Utara, Jawa Barat, Jawa Tengah, dan Jawa Timur, memiliki tingkat ketersediaan sarana kesehatan yang tinggi, sedangkan 14 provinsi lainnya memiliki tingkat sarana kesehatan yang sedang, sementara 16 provinsi lainnya masuk dalam kategori tingkat ketersediaan sarana kesehatan yang rendah [15].

Penelitian berikutnya [16] mengindikasikan bahwa dengan menggunakan pendekatan pengelompokan dua cluster, data berhasil dikelompokkan berdasarkan potensi yang dimilikinya. Cluster 1 tergolong dalam kategori potensi tinggi dengan rata-rata tingkat kecerahan sebesar 344.470K dan tingkat kepercayaan rata-rata sebesar 87.08%, sementara Cluster 2 masuk dalam kategori potensi sedang dengan rata-rata tingkat kecerahan sebesar 318.800K dan tingkat kepercayaan rata-rata sebesar 58.73%. Di sisi lain, penelitian yang dilakukan oleh [17] menghasilkan dataset dengan kode yang mencakup seluruh data yang mendapatkan predikat terbaik dalam hal keseragaman dalam proses pengelompokan dengan nilai 2,245. Dataset yang mencakup seluruh kode ini menunjukkan bahwa nilai koefisien keseragaman atau CCC (Concordance Correlation Coefficient) berkisar antara 2 hingga 3, mengindikasikan bahwa dataset tersebut memiliki tingkat keseragaman yang baik. Dari berbagai studi yang telah dilakukan, dapat disimpulkan bahwa data mining adalah serangkaian proses yang digunakan untuk menghasilkan informasi bernilai tambah dengan mengidentifikasi pola penting dari data yang tersimpan dalam basis data. Salah satu pendekatan dalam data mining adalah metode pengelompokkan hierarki, yang merupakan komponen dari pengelompokkan partitional. Metode ini mengambil objek tertentu dalam kumpulan objek untuk mewakili sebuah kelompok. "Keunggulan utama dari metode ini adalah kemampuannya untuk mengatasi kelemahan yang ada pada metode k-means yang

rentan terhadap data ekstrem dan fakta bahwa hasil pengelompokkan tidak dipengaruhi oleh urutan entri dalam dataset" [18].

3. METODE PENELITIAN

3.1. Analisa Permasalahan

Penelitian ini bertujuan untuk menerapkan teknik data mining dalam proses pengelompokkan data imunisasi guna mengidentifikasi daerah-daerah dengan tingkat imunisasi yang rendah, sehingga upaya sosialisasi dapat diarahkan kepada daerah-daerah tersebut. Hasil analisis yang disajikan dalam penelitian ini mengungkapkan tantangan mendasar yang dihadapi adalah peningkatan kasus penyakit campak pada balita yang terjadi setiap tahun. Indonesia secara internasional termasuk dalam 10 negara dengan jumlah kasus penyakit campak terbanyak.

Untuk memutuskan mata rantai penularan penyakit campak, khususnya pada balita, dibutuhkan tingkat imunisasi minimal sebesar 95%. Hal ini semakin penting mengingat bahwa tingkat cakupan imunisasi campak di Indonesia saat ini hanya mencapai 84% dan masih berada dalam kategori imunisasi campak yang sedang jika dibandingkan dengan negara-negara tetangga di Asia Tenggara. Oleh karena itu, dibutuhkan upaya lebih lanjut, termasuk usaha pemerataan imunisasi campak di seluruh wilayah Ibukota Jakarta, Indonesia.

3.2. Metode Pengumpulan Data

Dalam pengumpulan data, penulis bergantung pada beragam sumber referensi seperti perpustakaan, buku, prosiding, atau jurnal, untuk mendapatkan informasi yang digunakan dalam menentukan parameter-parameter dalam penelitian ini. Data penelitian diperoleh dari data.jakarta.go.id. Data yang digunakan dalam penelitian ini adalah data mengenai persentase imunisasi campak pada anak balita di berbagai wilayah di Ibukota Jakarta, yang diambil dari tahun 2019. Variabel yang menjadi fokus dalam penelitian adalah akumulasi persentase imunisasi.

3.3. Analisis Data

Proses analisis data dapat dilakukan setelah selesai mengumpulkan data. Dalam penelitian ini, penulis melakukan analisis statistik deskriptif. Jenis data yang digunakan adalah data sekunder. Data sekunder merujuk pada data yang tidak diperoleh langsung dari sumber aslinya, melainkan sudah dikumpulkan dan diolah sebelumnya oleh pihak lain, dan data ini memiliki relevansi dengan isu yang sedang diteliti. Data dalam studi ini diolah menggunakan aplikasi KNIME, KNIME adalah ilmu data kode rendah dan platform persiapan data yang membuat pemahaman data dan merancang alur kerja analitik dapat diakses oleh semua orang.

3.4. Algoritma Clustering Hirarki

Metode clustering hirarki, atau pengelompokan hierarkis adalah sebuah metode yang menghasilkan hierarki struktur atau tingkatan dari kumpulan data atau objek berdasarkan kesamaan karakteristiknya. Cluster yang diinginkan dalam metode ini biasanya belum diketahui jumlahnya, dan hasilnya sering kali direpresentasikan dalam bentuk dendrogram untuk mempermudah pemahaman hierarki tersebut. Penggunaan metode ini memiliki sejumlah keuntungan, seperti mempercepat proses pengolahan dan menghemat waktu karena keluaran berbentuk hierarki yang mudah untuk diinterpretasikan. Namun, ada beberapa kekurangan dalam metode ini, seperti kerentanannya terhadap data yang berbeda dari norma (outlier), variasi dalam skala jarak yang digunakan, dan variabel-variabel yang mungkin tidak relevan [19].

Metode ini memiliki dua pendekatan utama, yaitu pendekatan agglomeratif (penggabungan) dan pendekatan divisif (pemisahan). Pendekatan agglomeratif dimulai dengan memperlakukan setiap titik sebagai cluster individu, dan pada setiap langkah, pasangan titik dalam cluster digabungkan hingga hanya satu titik atau cluster tunggal yang tersisa [20]. Dalam pendekatan agglomeratif ini, terdapat lima metode yang digunakan, yakni average linkage (berdasarkan rata-rata jarak antara seluruh individu dalam suatu cluster dengan seluruh individu dalam cluster lain), complete linkage (berdasarkan jarak terjauh), single linkage (berdasarkan jarak terdekat), metode Ward (berdasarkan jumlah total kuadrat dua cluster pada setiap variabel), dan metode centroid (berdasarkan jarak antara titik pusat dua cluster yang berhubungan) [21][22].

Pendekatan divisive dimulai dengan satu cluster besar yang mencakup semua titik data. Pada setiap langkah dalam proses ini, cluster tersebut dibagi-bagi hingga setiap cluster hanya berisi satu titik data, bertolak dari cara yang berkebalikan dengan pendekatan agglomerative. Berikut adalah tahapan dalam melakukan analisis Clustering hierarki:

a. Mengukur tingkat kemiripan antara objek (similaritas). Sesuai dengan prinsip analisis clustering yang memfokuskan pada pengelompokan objek berdasarkan tingkat kemiripannya, tahap awal adalah mengukur sejauh mana objek-objek tersebut memiliki kesamaan.
b. Membentuk cluster dengan menggunakan metode yang telah ditetapkan.
c. Menganalisis hasil dari proses pengelompokan. Setelah pembentukan cluster, langkah selanjutnya adalah melakukan interpretasi terhadap cluster yang telah terbentuk, sehingga dapat menjelaskan karakteristik kumpulan data dalam masing-masing cluster.
d. Memvalidasi cluster. Tahap ini melibatkan pengujian cluster yang telah terbentuk. Selanjutnya, dilakukan proses profiling untuk

menggambarkan karakteristik masing-masing cluster berdasarkan profil tertentu, seperti usia konsumen pembeli rumah dan tingkat penghasilannya. Validasi diperlukan untuk memastikan bahwa hasil clustering merepresentasikan populasi secara keseluruhan dan dapat digeneralisasikan ke objek lainnya serta tetap stabil dalam periode waktu tertentu.

e. Asumsi dalam Analisis Cluster. Asumsi yang perlu dipenuhi mencakup representativitas sampel yang harus mewakili populasi yang lebih besar serta penanganan masalah multikolinearitas.

4. HASIL DAN PEMBAHASAN

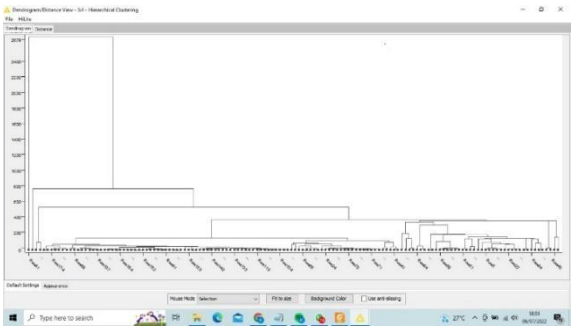
4.1. Perspektif Analisis Dan Pemrosesan Data

Berdasarkan ahli dalam bidang kesehatan yang membuat dataset ini, hasil positif dan negative dari vaksinasi dapat di tentukan dengan satu dimensi atau bahkan dengan mengkombinasikan beberapa dimensi. Namun, kenyataan yang dimiliki dataset berikut adalah hampir sebagian besar uji coba yang kami lakukan menghasilkan hasil positif dan negative yang tidak menentu berdasarkan sebuah dimensi. Perlu di ingat kembali dalam studi ini, kami menganalisis kumpulan data yang dijelaskan di atas melalui metode berbasis clsutering hierarki untuk menemukan berbagai alasan untuk hasil yang berbeda. Ide kami adalah untuk menerapkan algoritma clustering hierarki ke dataset untuk menemukan hasil immunisasi dari kelas hasil yang sama.

Seperti yang ditunjukkan di atas, kumpulan data yang dipertimbangkan adalah heterogen karena mengandung dimensi numerik dan nominal. Algoritma clustering hierarki yang ada biasanya bekerja pada data numerik saja. Juga, untuk implementasi yang ada tidak ada deskripsi tentang bagaimana nilai yang hilang diperlakukan. Sebagai konsekuensinya, kami melakukan praproses dataset dengan cara berikut: (1) Kami menghapus semua catatan yang memiliki nilai yang hilang di salah satu dimensinya. (2) Kami mengubah semua dimensi nominal seperti wilayah, kecamatan, atau jenis immunisasi menjadi representasi numerik. Karena kenyataan bahwa semua dimensi nominal hanya terdiri dari dua nilai yang berbeda (terutama ya dan tidak), kami mengonversi nilai menjadi 0 atau 1. Akhirnya, kami menormalkan semua dimensi linier dalam rentang [0, 1]. Setelah ini, semua dimensi numerik dalam kisaran [0, 1], memungkinkan analisis lebih lanjut dengan dimensi berbobot sama.

4.2. Analisis dari Proses dan Hasil Eksperimen

Dalam percobaan awal kami pada dataset, kami menemukan bahwa sebagian besar analisis tidak berguna. Kami menggunakan alat penambangan data seperti KNIME [18] untuk mengelompokkan pasien ke dalam kelompok yang berbeda, atau menerapkan algoritma klasifikasi yang berbeda untuk memprediksi dan mengelompokan hasil immunisasi dengan benar.



Gambar 3. Dendrogram yang mengilustrasikan pengelompokan hierarkis dari dataset studi ini.

Clustering hierarki diterapkan. Hasilnya diilustrasikan sebagai Dendrogram pada Gambar. 3. Sumbu x dipetakan ke masing-masing data, sumbu y mewakili perbedaan antara dua data atau kelompok data. Nilai y yang besar sesuai dengan ketidakmiripan yang tinggi. Dari Gambar. 3 kita melihat bahwa tidak ada data yang dianggap serupa dan, sebagai akibatnya, tidak ada pengelompokan data yang berguna yang dapat diidentifikasi. Kami mengasumsikan alasannya sebagai: (1) Data biasanya mirip satu sama lain hanya dalam subset dimensi, (2) kesamaan dalam satu dimensi dapat dilawan oleh perbedaan dalam dimensi lain, dan (3) efek konsentrasi mempengaruhi perhitungan kesamaan dalam ruang dimensi tinggi.

Untuk tugas klasifikasi, tidak menghapus nilai yang hilang melainkan menggantinya dengan nilai rata-rata dimensi. Kami menerapkan beberapa algoritma klasifikasi untuk menemukan prediktor yang berguna untuk hasil vaksinasi. Eksperimen kami terdiri dari, misalnya, Decision Tree, Naive Bayes, dan Random Forest. Kami membagi dataset menjadi satu set pelatihan (80% dari catatan) dan satu set validasi (20% dari catatan). Untuk validasi, kami mengukur persentase data yang diklasifikasikan dengan benar setelah pelatihan model. Sementara akurasi klasifikasi dataset pelatihan berkinerja sangat baik (sekitar 84% untuk random forest), akurasi untuk dataset validasi turun di bawah 50% untuk beberapa algoritma; yang lebih buruk dari klasifikasi acak. Kami berasumsi kinerja klasifikasi yang buruk disebabkan oleh (1) ukuran dataset pelatihan yang terlalu kecil, dan (2) tidak ada aspek global yang memungkinkan klasifikasi ke dalam dua kelas hasil. Sebaliknya, kombinasi fitur yang berbeda mungkin relevan untuk memprediksi hasil dengan benar.

5. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa Ilmubiomedis dan perawatan kesehatan berubah menjadi ilmu data intensif, di mana kita tidak hanya menghadapi peningkatan volume dan keragaman data yang sangat kompleks, multi-dimensi dan seringkali terstruktur

dengan lemah, tetapi juga kebutuhan yang berkembang untuk analisis integratif dan pemodelan. Mempertimbangkan bahwa analisis dalam ruang atribut (fitur) lengkap mungkin tidak efektif, kami di sini mengeksplorasi analisis klaster hierarki untuk mempelajari mengelompokan data hasil pengobatan imunisasi pada anak sekolah di daerah ibukota jakarta.

Kami menemukan bahwa segmentasi data untuk hasil pengobatan diikuti oleh pengelompokan subruang memungkinkan untuk mengidentifikasi kelompok data yang relevan, yang dapat membentuk dasar untuk generalisasi pengetahuan medis. Alur kerja yang kami usulkan hanyalah langkah pertama, dan kami mengidentifikasi sejumlah tantangan dan perluasan menarik untuk pekerjaan masa depan di area tersebut. Visi besar untuk masa depan adalah untuk secara efektif mendukung pembelajaran manusia dengan pembelajaran mesin - visualisasi dekat dengan pengguna akhir, karenanya sangat diperlukan dalam pendekatan ini.

DAFTAR PUSTAKA

[1] M. S. Sarfraz, N. Murray, V. Sharma, A. Diba, L. Van Gool, and R. Stiefelhagen, "Temporally-Weighted Hierarchical Clustering for Unsupervised Action Segmentation."

[2] C. H. Lee, C. H. Hung, and S. J. Lee, "A comparative study of divisive hierarchical clustering algorithms," *SNPD 2013 - 14th ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput.*, pp. 557–562, 2013, doi: 10.1109/SNPD.2013.6.

[3] K. S. Ashit Kumar Dutta, Mohamed Elhoseny, Vandna Dahiya, "An efficient hierarchical clustering protocol for multihop Internet of vehicles communication," 2019.

[4] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical Clustering with Hard-batch Triplet Loss for Person Re-identification."

[5] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data."

[6] and Z. B. M. Jafarzadegan, F. Safi-Esfahani, "Combining hierarchical clustering approaches using the PCA method," *Expert Syst. Appl*, vol. 137, pp. 1–10, 2019.

[7] M. T. Alessio Ishizaka, Banu Lokman, "A Stochastic Multi-criteria divisive hierarchical clustering algorithm," *Omega*, vol. 103, 2021.

[8] D. Ghoshdastidar, "Foundations of Comparison-Based Hierarchical Clustering."

[9] H. R. Y. & M. K. Subbulakshmi Pasupathi, Vimal Shanmuganathan, Kaliappan Madasamy, "Trend analysis using agglomerative hierarchical clustering approach for time series big data," *J. Supercomput*, vol. 77, pp. 6505–6524, 2021.

[10] and J. M. K. Li, Z. Ma, D. Robinson, "Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering," *Appl. Energy*, vol. 231, pp. 331–342, 2018.

[11] S. Dalimunthe and A. Hanafiah, "Implementation of Agglomerative Hierarchical Clustering Based on The Classification of Food Ingredients Content of Nutritional Substances," vol. 6, no. 1, pp. 60–69, 2021.

[12] W. Paper, "Using Hierarchical Clustering Algorithms for Turkish Residential Market," 2012, doi: 10.5539/ijef.v4n1p138.

[13] I. Indeks and P. Manusia, "PENERAPAN ANALISIS CLUSTER DENGAN METODE HIERARKI UNTUK KLASIFIKASI KABUPATEN/KOTA DI PROVINSI MALUKU BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA," vol. 2, no. 2, pp. 123–130, 2020.

[14] Y. Rani and H. Rohil, "A Study of Hierarchical Clustering Algorithm," vol. 3, no. 11, pp. 1225–1232, 2013.

[15] G. Carlsson, "Characterization , Stability and Convergence of Hierarchical," vol. 11, pp. 1425–1470, 2010.

[16] D. T. Utari and D. S. Hanun, "Hierarchical Clustering Approach for Region Analysis of Contraceptive Users," vol. 2, no. 2, pp. 99–108, 2021, doi: 10.20885/EKSAKTA.vol2.iss2.art3.

[17] W. Laksito, Y. Saptomo, Y. Retno, and W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," no. 1, pp. 75–87, 2020.

[18] C. C. Astuti and R. S. Untari, "Applied Hierarchical Cluster Analysis with Average Linkage Algoritm," vol. 5, no. January, pp. 1–7, 2017.

[19] M. Thu and M. Moe, "A Modified Hierarchical Agglomerative Approach for Efficient Document Clustering System," pp. 228–238.

[20] "A Review on Hierarchiecal Clustering Algorithms." .

[21] O. Access, "Comparison of hierarchical cluster analysis methods by cophenetic correlation," pp. 1–8, 2013.

[22] V. Cohen-addad and S. Université, "Hierarchical Clustering : Objective Functions and Algorithms," vol. 66, no. 4