

Analisis Pola Penyakit Pasien Di Puskesmas Jayengan Surakarta Menggunakan Algoritma *K-Prototype*

Siti Solikah¹⁾, Sri Wulandari²⁾, Resia Perwirani³⁾

^{1,2,3} Politeknik Indonusa Surakarta

^{1,2,3} JL. KH. Samanhudi No.31, Bumi Laweyan, Surakarta

¹f22100@poltekindonusa.ac.id, ²sriwulandari@poltekindonusa.ac.id,

³resiaperwirani@gmail.com

Abstrak

Tingginya kunjungan pasien di Puskesmas menunjukkan pentingnya pemanfaatan data rekam medis untuk mengidentifikasi pola penyakit secara menyeluruh. Penelitian ini bertujuan menganalisis pola penyakit pasien di Puskesmas Jayengan Surakarta menggunakan algoritma *K-Prototype* dengan variabel usia, jenis kelamin, diagnosis, dan alamat melalui pendekatan *Knowledge Discovery in Database* (KDD). Hasil menunjukkan bahwa diagnosis tertinggi adalah *General examination and investigation of persons without complaint and reported diagnosis* (3.375 kasus), diikuti *Acute nasopharyngitis* (2.896 kasus) dan *Essential (primary) hypertension* (1.855 kasus), dengan wilayah Tipes dan Serengan serta pasien perempuan (63,89%) sebagai dominan. *Clustering* menghasilkan dua klaster optimal dengan nilai DBI sebesar 7,13 dan Silhouette sebesar 0,148, dimana klaster 1 didominasi usia dewasa-lansia dengan kecenderungan penyakit degeneratif dan klaster 2 didominasi usia anak-remaja dengan kecenderungan penyakit infeksi. Metode *K-Prototype* terbukti efektif dan dapat mendukung perencanaan program kesehatan berbasis karakteristik pasien.

Kata kunci: Data mining, *Clustering*, *K-Prototype*, Rekam Medis

Abstract

The high number of patient visits at community health centers highlights the importance of utilizing medical record data to identify disease patterns comprehensively. This study aims to analyze patient disease patterns at Jayengan Public Health Center, Surakarta, using the *K-Prototype* algorithm with variables including age, gender, diagnosis, and address through the *Knowledge Discovery in Database* (KDD) approach. The results show that the highest diagnosis was *General examination and investigation of persons without complaint and reported diagnosis* (3,375 cases), followed by *Acute nasopharyngitis* (2,896 cases) and *Essential (primary) hypertension* (1,855 cases), with Tipes and Serengan areas and female patients (63.89%) being dominant. The clustering process produced two optimal clusters with a Davies-Bouldin Index of 7.13 and a Silhouette Score of 0.148, where cluster 1 was dominated by adults and the elderly with a tendency toward degenerative diseases, while cluster 2 was dominated by children and adolescents with a tendency toward infectious diseases. The *K-Prototype* method proved effective and can support health program planning based on patient characteristics.

Keywords: Data mining, Clustering, *K-Prototype*, Medical Records

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong transformasi di sektor kesehatan melalui pemanfaatan teknologi informasi kesehatan dalam pengelolaan data secara cepat, akurat, dan terintegrasi (Fauzi et al., 2024). Penerapan rekam medis elektronik di puskesmas menjadi upaya penting dalam meningkatkan mutu layanan serta efisiensi administrasi. Sebagai fasilitas kesehatan tingkat pertama, puskesmas memiliki peran strategis dalam menyediakan layanan berbasis data guna mendukung peningkatan derajat kesehatan masyarakat (Panggabean et al., 2025).

Namun demikian, tingginya beban penyakit di Indonesia serta besarnya volume kunjungan pasien di Puskesmas Jayengan Surakarta menunjukkan bahwa pengelolaan data kesehatan masih belum optimal. Sistem pengelompokan penyakit yang digunakan saat ini masih dilakukan secara terpisah berdasarkan program, sehingga belum mampu mengidentifikasi pola penyakit secara menyeluruh dan terintegrasi. Kondisi ini menyebabkan keterbatasan dalam pemanfaatan data sebagai dasar pengambilan keputusan dan perencanaan layanan kesehatan tepat sasaran.

Pendekatan data mining, khususnya teknik *clustering*, dapat digunakan untuk mengatasi permasalahan tersebut dengan mengelompokkan data berdasarkan kesamaan karakteristik. Algoritma *K-Means* merupakan metode yang umum digunakan, namun memiliki keterbatasan karena hanya dapat mengolah data numerik (Putra & Fadhillah, 2025). Sementara itu, data kesehatan memiliki karakteristik campuran, seperti usia (numerik) serta jenis kelamin, diagnosis, dan alamat (kategorik). Oleh karena itu, algoritma *K-Prototype* menjadi alternatif yang lebih sesuai karena mampu mengolah data campuran secara langsung (Nooraeni, 2023).

Beberapa penelitian terdahulu menunjukkan bahwa metode *clustering* efektif dalam mengidentifikasi pola penyakit pasien. Penelitian oleh (Fajri & Purnamasari, 2022) menggunakan algoritma *K-Means* dengan variabel usia, jenis kelamin, dan diagnosis, menghasilkan tiga klaster dengan dominasi penderita pada kelompok usia dewasa dan anak-anak serta jenis penyakit seperti asma, diare, hipertensi, dan infeksi saluran pernapasan. Penelitian oleh (Harmaja et al., 2023) juga menerapkan *K-Means* dengan

variabel yang sama dan menghasilkan tiga klaster, dimana klaster tertinggi didominasi oleh pasien laki-laki berusia diatas 40 tahun dengan penyakit infeksi saluran pernapasan akut (ISPA). Selain itu, penelitian (Laksono et al., 2024) menunjukkan bahwa pengelompokan data pasien berdasarkan usia, jenis kelamin, dan diagnosis mampu menghasilkan beberapa klaster yang menggambarkan keterkaitan antara kelompok usia dengan jenis penyakit yang dominan. Namun, penelitian-penelitian tersebut masih memiliki keterbatasan, yaitu penggunaan algoritma *K-Means* yang tidak mampu menangani data campuran serta belum mempertimbangkan variabel spasial seperti alamat pasien.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis pola penyakit pasien di Puskesmas Jayengan Surakarta menggunakan algoritma *K-Prototype* dengan mempertimbangkan variabel usia, jenis kelamin, diagnosis, dan alamat. Penelitian ini diharapkan dapat menghasilkan informasi pola penyakit yang lebih komperhensif dan terintegrasi sebagai dasar pengambilan keputusan layanan kesehatan.

2. TINJAUAN PUSTAKA

a. Puskesmas

Puskesmas merupakan fasilitas pelayanan kesehatan tingkat pertama yang bertanggung jawab dalam penyelenggaraan pelayanan kesehatan promotif, preventif, kuratif, rehabilitatif, dan paliatif dengan penekanan pada upaya promotif dan preventif. Puskesmas memiliki wilayah kerja tertentu dan berperan sebagai garda terdepan dalam memberikan pelayanan kesehatan yang merata, terjangkau, dan berkesinambungan kepada masyarakat. Selain itu, puskesmas berfungsi dalam meningkatkan derajat kesehatan pada setiap tahap kehidupan, perbaikan faktor determinan kesehatan, serta penguatan kapasitas individu, keluarga, dan masyarakat (Permenkes RI, 2024).

b. Rekam Medis

Rekam medis merupakan dokumen yang berisi informasi identitas pasien, hasil pemeriksaan, diagnosis, pengobatan, serta tindakan medis yang diberikan. Rekam medis memiliki peran penting dalam mendukung proses pelayanan kesehatan, baik sebagai dasar pengambilan keputusan klinis maupun sebagai sumber data untuk

analisis kesehatan. Seiring perkembangan teknologi, rekam medis kini dikembangkan dalam bentuk elektronik yang memungkinkan pengelolaan data secara lebih sistematis, efisien, dan terintegrasi (Permenkes RI, 2022). Data dalam rekam medis terdiri atas data administratif dan data klinis yang dapat dimanfaatkan untuk analisis tren penyakit serta perencanaan layanan kesehatan (Nurchayati & Lestari, 2022).

c. Data Mining

Data mining merupakan suatu proses menemukan pengetahuan atau pola tersembunyi dari data dalam jumlah besar untuk menghasilkan informasi yang bermanfaat dalam pengambilan keputusan (Amna et al., 2023). Secara umum, data mining terbagi menjadi dua jenis, yaitu deskriptif yang bertujuan menemukan pola dalam data, dan prediktif yang digunakan untuk memperkirakan nilai dimasa mendatang (Pebdika et al., 2023).

Teknik dalam data mining meliputi klasifikasi, regresi, *clustering*, asosiasi, dan analisis anomali (Hapsari et al., 2024). Proses data mining merupakan bagian dari *Knowledge Discovery in Database* (KDD) yang meliputi tahapan seleksi data, *preprocessing*, transformasi, data mining, serta interpretasi hasil (Prasetyani & Rochmawati, 2022).

d. Clustering

Clustering merupakan teknik dalam data mining yang digunakan untuk mengelompokkan data kedalam beberapa klaster berdasarkan kesamaan karakteristik tanpa adanya label sebelumnya. Tujuan utama *clustering* adalah menemukan pola atau struktur tersembunyi dalam data (Hendrastuty, 2024). Metode *clustering* memiliki berbagai pendekatan, seperti *partitioning*, *hierarkis*, berbasis densitas, berbasis model, dan berbasis *grid*. Pemilihan metode *clustering* disesuaikan dengan karakteristik data dan tujuan analisis (Sudipa et al., 2024).

e. Algoritma *K-Prototype*

K-Prototype merupakan algoritma *clustering* berbasis *partitioning* yang dikembangkan untuk mengatasi keterbatasan *K-Means* dalam menangani data bertipe campuran (Fadilah & Wijayanto, 2023). Algoritma ini menggabungkan perhitungan jarak *euclidean* untuk data numerik dan metode *K-Modes* untuk data kategorikal dalam

menentukan kedekatan antar data. Proses *clustering* dilakukan melalui penentuan jumlah klaster, inisialisasi pusat klaster, perhitungan jarak, pengelompokan data, serta pembaruan pusat klaster hingga mencapai kondisi *konvergen* (Purba et al., 2024). Keunggulan utama *K-Prototype* adalah kemampuannya mengolah data numerik dan kategorikal secara bersamaan dengan kompleksitas yang relatif rendah, sehingga efektif digunakan pada data berskala besar seperti data rekam medis (Wijayati & Saputro, 2023).

Pemanfaatan algoritma *K-Prototype* pada data rekam medis yang mencakup data administratif dan klinis memungkinkan indentifikasi pola penyakit secara lebih komperhensif, sehingga dapat mendukung pengambilan keputusan dalam pelayanan kesehatan (Purnomo et al., 2025).

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan teknik *clustering* untuk mengidentifikasi pola penyakit pasien berdasarkan data sekunder yang diperoleh dari rekam medis elektronik di Puskesmas Jayengan Surakarta periode januari- desember 2024. Variabel yang digunakan meliputi usia, jenis kelamin, diagnosis, dan alamat. Metode yang diterapkan adalah algoritma *K-Prototype*, yaitu metode *clustering* berbasis *partitioning* yang mampu mengolah data bertipe campuran (numerik dan kategorikal) sesuai dengan karakteristik data rekam medis. Sampel penelitian diperoleh menggunakan teknik *purposive sampling* dengan kriteria data pasien yang memiliki atribut lengkap dan termasuk pasien rawat jalan, sementara data duplikat dan pasien rujukan dikeluarkan dari analisis.

Analisis data dalam penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahapan sebagai berikut:

a. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data sekunder yang berasal dari rekam medis elektronik periode 2024 dalam bentuk format excel.

b. Data Selection (Pemilihan Data)

Tahap ini dilakukan untuk memilih variabel yang relevan dengan tujuan penelitian, yaitu usia, jenis kelamin, diagnosis, dan alamat. Selain itu, dilakukan pemilihan sampel berdasarkan kreteria inklusi dan

eksklusi yang telah ditetapkan, sehingga data yang digunakan benar – benar sesuai dengan kebutuhan analisis.

c. *Preprocessing*

Tahap *preprocessing* bertujuan untuk memastikan kualitas data sebelum dilakukan analisis. Proses yang dilakukan meliputi penghapusan data duplikat, penanganan data kosong atau tidak lengkap dengan cara menghapus atau mengganti nilai tertentu, serta deteksi *outlier* untuk mengidentifikasi data yang menyimpang dari pola umum.

d. *Data mining*

Pada tahap ini dilakukan proses *clustering* menggunakan algoritma *K-Prototype* untuk mengelompokkan data berdasarkan kesamaan karakteristik. Proses pengolahan data dilakukan menggunakan bahasa pemrograman *python* melalui *gogle colab*, termasuk dalam menentukan jumlah klaster yang belum optimal. Selain itu, *Power BI* digunakan untuk menampilkan hasil dalam bentuk visualisasi seperti grafik dan diagram guna mempermudah pemahaman pola data.

e. *Evaluation*

Evaluasi dilakukan untuk menilai kualitas hasil pengelompokan yang terbentuk. Metode yang digunakan adalah *Davies -Bouldin Index (DBI)*, yang mengukur tingkat kedekatan antar data dalam satu klaster serta tingkat pemisahan antar klaster, sehingga dapat menunjukkan seberapa baik kualitas klaster yang dihasilkan.

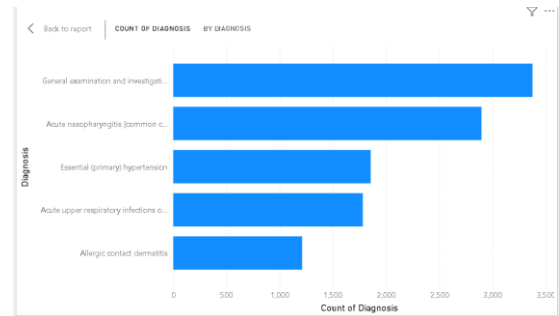
f. *Interpretasi*

Tahap akhir adalah interpretasi hasil *clustering* dengan menganalisis karakteristik masing-masing klaster. Pada tahap ini diidentifikasi pola, ciri khas, serta kecenderungan penyakit pada setiap kelompok, sehingga dapat memberikan gambaran yang lebih komperhensif terkait kondisi kesehatan pasien.

4. HASIL DAN PEMBAHASAN

Data penelitian diperoleh dari rekam medis pasien di Puskesmas Jayengan Surakarta tahun 2024 dengan jumlah awal 62.728 data dan 48 atribut. Setelah dilakukan proses seleksi data, pembersihan data duplikasi, penanganan *missing value*, serta deteksi *outlier*, jumlah

data yang digunakan dalam analisis menjadi 31.939 data. Data yang telah bersih kemudian divisualisasikan menggunakan *Power BI* untuk memperoleh gambaran awal pola distribusi data.



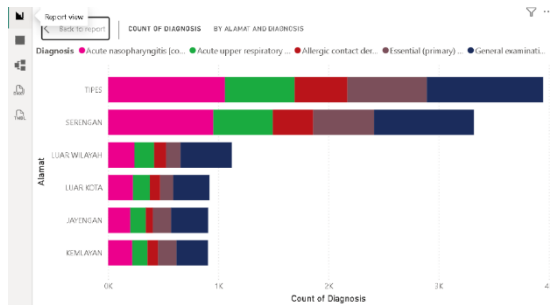
Gambar 1. Visualisasi 5 Diagnosis Tertinggi

Dari gambar diatas, hasil visualisasi menunjukkan bahwa diagnosis *General examination and investigation of persons without complaint and reported diagnosis* merupakan kasus tertinggi dengan jumlah 3.375 kasus. Hal ini menunjukkan bahwa pemeriksaan kesehatan umum banyak dimanfaatkan masyarakat sebagai upaya deteksi dini dan pemantauan kondisi kesehatan. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa tingginya kunjungan pada layanan pemeriksaan umum dipengaruhi oleh kebutuhan masyarakat terhadap pelayanan kesehatan kesehatan preventif serta kualitas pelayanan tenaga kesehatan (Dapa et al., 2024).

Diagnosis terbanyak berikutnya adalah *Acute nasopharyngitis (common cold)* sebanyak 2.896 kasus. Penyakit ini merupakan infeksi saluran pernapasan atas yang disebabkan oleh virus dengan gejala seperti demam, sakit kepala, dan gangguan tenggorokan (Rachel, 2025). Tingginya kasus ini menunjukkan bahwa infeksi saluran pernapasan masih menjadi masalah kesehatan yang dominan dimasyarakat. Hal ini juga dipengaruhi oleh faktor lingkungan, kepadatan penduduk, serta perilaku masyarakat seperti kebiasaan berkumpul di ruang tertutup yang dapat mempercepat penularan penyakit (Ritonga et al., 2025).

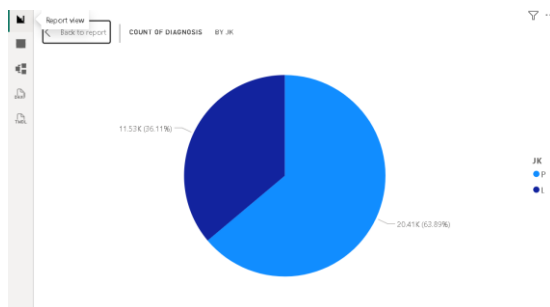
Selain itu, *Essential (primary) hypertension* termasuk dalam tiga besar diagnosis dengan jumlah 1.855 kasus. Hipertensi merupakan penyakit degeneratif yang ditandai dengan peningkatan tekanan darah $\geq 140/90$ mmHg dan sering disebut *silent killer* karena berkembang tanpa gejala namun berdampak serius (Lubis et al., 2023). Faktor

risiko hipertensi meliputi faktor yang tidak dapat dimodifikasi seperti usia dan genetik, serta faktor yang dapat dimodifikasi seperti pola makan, aktivitas fisik, dan gaya hidup (Putri et al., 2023).



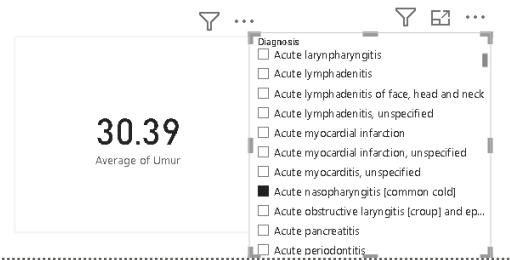
Gambar 2. Visualisasi Wilayah

Berdasarkan visualisasi wilayah, diketahui bahwa daerah Tipes dan Serangan memiliki jumlah kasus tertinggi yang didominasi oleh penyakit *Acute nasopharyngitis*. Hal ini dapat dipengaruhi oleh kepadatan penduduk yang tinggi sehingga meningkatkan risiko penularan penyakit. Sementara itu, wilayah Jayengan dan Kemlayan memiliki jumlah kasus yang relatif lebih rendah, yang kemungkinan dipengaruhi oleh kondisi demografis, akses terhadap fasilitas kesehatan, serta tingkat pemanfaatan layanan kesehatan oleh masyarakat (Amaliah et al., 2023).



Gambar 3. Visualisasi Distribusi Jenis Kelamin

Distribusi berdasarkan jenis kelamin menunjukkan bahwa pasien perempuan lebih banyak dibandingkan laki-laki, yaitu sebesar 63,89% berbanding 36,11%. Hal ini menunjukkan bahwa perempuan cenderung lebih aktif dalam memanfaatkan layanan kesehatan dibandingkan laki-laki. Penelitian sebelumnya juga menyebutkan bahwa perempuan lebih sering melakukan pemeriksaan kesehatan, sedangkan laki-laki cenderung datang ketika sudah mengalami keluhan (Dapa et al., 2024). Selain itu, faktor sosial dan budaya juga mempengaruhi perbedaan pemanfaatan layanan kesehatan antar *gender* (Sukiyem et al., 2025).



Gambar 4. Rata-rata Umur Berdasarkan Diagnosis

Hasil visualisasi rata-rata umur menunjukkan bahwa pasien dengan diagnosis *Acute nasopharyngitis* memiliki rata-rata umur sekitar 30,39 tahun. Penyakit ini termasuk dalam kelompok infeksi saluran pernapasan akut (ISPA) yang dapat menyerang berbagai kelompok usia, namun cukup banyak ditemukan pada usia dewasa karena tingginya aktivitas di luar ruangan serta paparan polusi dan lingkungan kerja (Rini et al., 2025).

Analisis *clustering* menggunakan algoritma *K-Prototype* menghasilkan dua kluster optimal berdasarkan evaluasi *Davies-Bouldin Index* dan *Silhouette Score*. Metode *K-Prototype* merupakan pengembangan dari *K-Means* yang mampu mengelompokkan data dengan atribut campuran, yaitu numerik dan kategorik (Fadilah & Wijayanto, 2023). Nilai DBI terendah (7,135) dan *Silhouette Score* tertinggi sebanyak (0,14890) diperoleh pada jumlah kluster 2, sehingga dianggap sebagai jumlah kluster terbaik dalam penelitian ini. Penggunaan metode evaluasi seperti DBI dan *Silhouette score* juga dilakukan pada penelitian sebelumnya untuk menentukan jumlah kluster optimal (Sholeh & Aeni, 2023).

Tabel 1. Distribusi Cluster

Cluster	
1	17291
2	14648

Berdasarkan tabel diatas kluster 1 terdiri dari 17.291 pasien dan didominasi oleh kelompok usia dewasa hingga lansia dengan rata-rata umur 56 tahun. Diagnosis yang dominan pada kluster ini adalah *Essential (primary) hypertension*, *Acute nasopharyngitis*, serta *General examination and investigation of persons without complaint and reported diagnosis*. Hal ini menunjukkan bahwa kluster 1 merepresentasikan kelompok penyakit degeneratif yang berkaitan dengan proses penuaan, yaitu akibat penurunan fungsi organ tubuh (Setyawati et al., 2024). Hipertensi pada kelompok ini dipengaruhi oleh perubahan fisiologis seperti penurunan elastisitas

pembuluh darah dan peningkatan resistensi perifer (Riyada et al., 2024), serta faktor gaya hidup seperti pola makan dan aktivitas fisik. Tingginya kasus *General examination and investigation of persons without complaint and reported diagnosis* menunjukkan adanya kesadaran untuk melakukan pemeriksaan rutin, sementara keberadaan *Acute nasopharyngitis* mengindikasikan bahwa kelompok ini tetap rentan terhadap infeksi akibat penurunan sistem imun.

Sementara itu, klaster 2 menunjukkan karakteristik penyakit infeksi dengan diagnosis *Acute nasopharyngitis*, *General examination and investigation of persons without complaint and reported diagnosis*, serta *Acute upper respiratory infection* pada beberapa bagian saluran pernapasan. Rata-rata umur pasien pada klaster ini sekitar 16 tahun, sehingga didominasi oleh kelompok usia anak hingga remaja. Hal ini menunjukkan bahwa klaster 2 berkaitan dengan penyakit infeksi, yaitu penyakit yang disebabkan oleh mikroorganisme patogen seperti bakteri, virus, parasit, maupun jamur yang dapat menular antar individu (Yulnefia & Sutia, 2022). Penyakit infeksi saluran pernapasan lebih sering terjadi pada kelompok usia ini karena faktor lingkungan, aktivitas sosial yang tinggi, serta tingkat penularan yang relatif mudah (Rini et al., 2025). Oleh karena itu, diperlukan upaya pencegahan melalui penerapan pola hidup bersih dan sehat, seperti pola makan bergizi, pemeriksaan kesehatan secara rutin, serta penggunaan masker saat berada diluar rumah untuk mengurangi paparan debu dan polusi (Aqida et al., 2024).

Pada kedua klaster ini, pasien perempuan mendominasi dibandingkan dengan pasien laki-laki, serta sebagian besar pasien berasal dari wilayah Tipes dan Serengan. Hal ini menunjukkan bahwa faktor demografis, kepadatan penduduk, dan perilaku masyarakat berperan penting dalam pola kunjungan pasien dan penyebaran penyakit, dimana wilayah dengan jumlah penduduk lebih padat cenderung memiliki jumlah kasus yang lebih tinggi serta tingkat pemanfaatan layanan kesehatan yang lebih besar.

Secara keseluruhan, metode *K-Prototype* mampu mengelompokkan data pasien secara efektif berdasarkan karakteristik usia, diagnosis, jenis kelamin, dan wilayah. Hasil pengelompokkan ini menghasilkan dua pola utama, yaitu kelompok penyakit degeneratif yang didominasi usia dewasa hingga lansia dan kelompok penyakit infeksi

yang didominasi usia anak hingga remaja. Temuan tersebut dapat dimanfaatkan sebagai dasar dalam perencanaan program kesehatan, seperti peningkatan skrining penyakit tidak menular pada kelompok usia lanjut serta upaya edukasi pencegahan infeksi pada kelompok usia muda.

5. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa diagnosis dengan kasus tertinggi adalah *General examination and investigation of persons without complaint and reported diagnosis* sebanyak 3.375 kasus, diikuti *Acute nasopharyngitis (common cold)* sebanyak 2.896 kasus, dan *Essential (primary) hypertension* sebanyak 1.855 kasus. Wilayah Tipes dan Serengan menjadi daerah dengan jumlah kasus tertinggi, dengan mayoritas pasien berjenis kelamin perempuan (63,89%) dibandingkan laki-laki (36,11%), serta rata-rata umur pasien *Acute nasopharyngitis* sekitar 30,39 tahun. Hasil analisis *clustering* menggunakan metode *K-Prototype* menghasilkan 2 klaster, yaitu klaster 1 sebanyak 17.291 pasien yang didominasi usia dewasa hingga lansia dengan rata-rata umur 56,08 tahun, serta klaster 2 sebanyak 14.648 pasien yang didominasi usia anak dan remaja dengan rata-rata umur 16,68 tahun. Kedua klaster menunjukkan dominasi pasien perempuan dan sebagian besar berasal dari wilayah Tipes dan Serengan. Berdasarkan evaluasi menggunakan *Davies-Bouldin Index* dan *Silhouette Score*, jumlah klaster optimal adalah 2 klaster dengan nilai DBI sebesar 7,13 dan *Silhouette Score* sebesar 0,148. Karakteristik klaster menunjukkan bahwa klaster 1 cenderung menggambarkan penyakit degeneratif dengan diagnosis dominan *Essential (primary) hypertension*, *Acute nasopharyngitis*, dan *General examination and investigation of persons without complaint and reported diagnosis*, sedangkan klaster 2 menggambarkan penyakit infeksi dengan diagnosis dominan *Acute nasopharyngitis*, *General examination and investigation of persons without complaint and reported diagnosis*, *Acute upper infection of multiple and unspecified sites*. Oleh karena itu, diperlukan upaya peningkatan pencegahan penyakit melalui edukasi perilaku hidup bersih dan sehat, khususnya untuk menekan kasus infeksi saluran pernapasan, serta peningkatan pemanfaatan layanan kesehatan pada laki-laki

melalui kegiatan skrining ditempat kerja, lingkungan masyarakat, dan fasilitas olahraga. Selain itu, wilayah Tipe dan Serangan perlu menjadi prioritas dalam program promosi kesehatan karena tingginya jumlah kunjungan pasien. Penelitian selanjutnya diharapkan dapat mengembangkan metode analisis dengan pendekatan lain pada data bertipe campuran, seperti *One-Hot Encoding* dan *Gower Distance*, guna meningkatkan kualitas hasil pengelompokan data.

6. REFERENSI

- Amaliah, H. A., Rizqiani, A., & Andayanie, E. (2023). Faktor Yang Berhubungan Dengan Minat Berkunjung Kembali Pada Pasien Di RSIA Permata Hati Makassar. *Window Of Public Health Journal*, 4(6), 968–976.
- Amna, S. W., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). Data Mining.
- Aqida, D. F., Sa'ban, Z., Syahrani, Z. P. A., Malolo, H. A., Utami, A., Rajab, B. N. A., & Ayu, R. D. (2024). Analisis Prioritas Masalah Kesehatan di Desa Bara Batu Kabupaten Pangkajene dan Kepulauan. *Jurnal Pengabdian Kepada Masyarakat*, 5, 14–23.
- Dapa, A., Kurniyanti, M. A., & Retnaningrum, D. N. (2024). Gambaran Pemanfaatan Fasilitas Kesehatan Di Puskesmas Mojolangu Malang. *Media Husada Journal of Nursing Science*, 5(3).
- Fadilah, Z. R., & Wijayanto, A. W. (2023). Perbandingan Metode Klasterisasi Data Bertipe Campuran : *One-Hot- Encoding*, *Gower Distance*, dan *K-Prototype* Berdasarkan Akurasi (Studi Kasus : Chronic Kidney Disease Dataset). *Journal of Applied Informatics and Computing (JAIC)*, 7(1).
- Fajri, M. B., & Purnamasari, S. D. (2022). Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan *K-Means Clustering*. *Journal of Information Technology Ampera*, 3(3), 317–334.
- Fauzi, M. R., Saimi, & Fathoni. (2024). Tantangan dan Solusi Administrasi Kesehatan di Era Digital (Tinjauan Literature Review atas Implementasi Teknologi). *Jurnal Studi Islam Dan Humaniora*, 5(1), 1093–1103.
- Hapsari, R. K., Sumarni, T., Hidayatulloh, T., Farida, Herianto, Elfaladonna, F., Aini, N., Nurfitriya, Kurniawan, H., Kraugusteeliana, Yusuf, L., & Widiatama, Y. (2024). Data Mining.
- Harmaja, O. J., Halawa, H., Hulu, W. S., & Loi, S. (2023). Implementasi Algoritma *K-Means Clustering* Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan. *Sains Dan Teknologi*, 5(1), 150–157.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma *K-Means Clustering* Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Laksono, B. W., Syahidin, Y., & Yunengsih, Y. (2024). Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap dengan Algoritma *K-Means Clustering*. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(2), 621–627. <https://doi.org/10.32493/jtsi.v7i2.39354>
- Lubis, A. F., Hutabarat, M. S. H., Aeni, M. N., & Sudarmi, L. (2023). Skrining Penyakit Degeneratif Pada Usia Produktif di Kelurahan Sukarami Sumatera Selatan. *Jurnal Pengabdian Kepada Masyarakat Nusantara*, 4(4), 5173–5178.
- Nooraeni, R. (2023). Metode *Cluster* Menggunakan Kombinasi Algoritma *Cluster K-Prototype* Dan Algoritma Genetika Untuk Data Bertipe Campuran. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 7(2), 81–97.
- Nurcahyati, S., & Lestari, N. D. (2022). Pemanfaatan Data Rekam Medis Dalam Pelaporan Bulanan Di Puskesmas Kejaksan Cirebon. *Indonesian Journal of Health Information Management*, 2(1), 1–5. <https://doi.org/10.54877/ijhim.v2i1.40>
- Panggabean, A. D. M., Suryandari, P. I., Bisono, E. F., & Ardila, N. M. I. (2025). Transformasi Digital Puskesmas: Meningkatkan Efektivitas Melalui Manajemen Rekam Medis Elektronik. *Indonesia Berdaya*, 6(2), 539–544.
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode *Naive Bayes* Untuk Menentukan Calon Penerima Pip. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1),

- 452–458.
<https://doi.org/10.36040/jati.v7i1.6303>
- Permenkes RI. (2022). Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 Tentang Rekam Medis. 33(1), 1–12.
- Permenkes RI. (2024). Peraturan Menteri Kesehatan Republik Indonesia Nomor 19 Tahun 2024 Tentang Penyelenggaraan Pusat Kesehatan Masyarakat.
- Prasetyani, S. D., & Rochmawati, N. (2022). Penerapan Data Mining Untuk *Clustering* Menu Favorit Menggunakan Algoritma *K-Means* (Studi Kasus Kedai Expo). *Journal of Informatics and Computer Science (JINACS)*, 3(03), 278–286. <https://doi.org/10.26740/jinacs.v3n03.p278-286>
- Purba, A. F., Mustafid, & Kartikasari, P. (2024). Penerapan Algoritma *K-Prototype* Untuk Pengelompokan Desa Di Kabupaten Bekasi Berdasarkan Infrastruktur Digital. *Jurnal Gaussian*, 13(2), 479–489. <https://doi.org/10.14710/j.gauss.13.2.479-489>
- Purnomo, H., Pambudi, R. E., & Irawan, R. (2025). Penerapan *Decision Tree* untuk Klasifikasi Penyakit Berdasarkan Data Rekam Medis. *Aisyah Journal of Informatics and Electrical Engineering*, 7(1), 1–8.
- Putra, I. P., & Fadhillah, A. (2025). Perbandingan Metode *K-Means* dan *Hierarchical Clustering* dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur. *Jurnal Inovasi Dan Tren*, 3(1), 227–234.
- Putri, L. M., Mamesah, M. M., & Sulistyana, C. S. (2023). Faktor Risiko Hipertensi Pada Masyarakat Usia Dewasa & Lansia Di Tambaksari Surabaya. *Journal Of Health Management Research*, 2(1), 1–6.
- Rachel. (2025). Pengertian Acute Nasopharyngitis. <https://id.abchealthonline.com/6586756-nasopharyngitis-symptoms-treatment-causes>
- Rini, D. K., Triana, N. Y., & Apriliyani, I. (2025). Gambaran Kejadian ISPA Di Puskesmas Purwojati. *Jurnal Ilmu Kesehatan*, 20(4), 2–7. <https://doi.org/10.5455/mnj.v1i2.644xa>
- Ritonga, F. R., Anjaini, N., & Gurning, F. P. (2025). Analisis *Trend* Penyakit Pada Peserta Jaminan Kesehatan Nasional Di Puskesmas Gunung Tinggi. *Jurnal Kolaboratif Sains*, 8(7), 4549–4555. <https://doi.org/10.56338/jks.v8i7.8239>
- Riyada, F., Fauziah, S. A., Liana, N., & Hasni, D. (2024). Faktor yang Mempengaruhi Terjadinya Resiko Hipertensi pada Lansia. *Scientific Journal*, 3(1).
- Setyawati, T., Putrie, I. R., Walanda, R. M., Oktafiani, D., Listawati, Zahid, S. R., Saleh, A. K. I., & Esa, M. B. (2024). Pemahaman Penyakit Degeneratif Pada Masyarakat Di Kelurahan Talise. *Jurnal Kolaboratif Sains*, 7(8), 3374–3383. <https://doi.org/10.56338/jks.v7i8.5928>
- Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode *Davies Bouldin*, *Elbow* Dan *Silhouette* Pada Model *Clustering* Dengan Menggunakan Algoritma *K-Means*. *Satuan Tulisan Riset Dan Inovasi Teknologi*, 8(1).
- Sudipa, I. G. I., Darmawiguna, I. G. M., Dendi, I. M., & ... (2024). Buku Ajar Data Mining. *PT* (Issue April).
- Sukiyem, Novalinda, C., & Ginting, J. B. (2025). Analisis Determinan Pemanfaatan Layanan Kesehatan Di Puskesmas Negeri Lama Labuhan batu Tahun 2024. *Jurnal Kedokteran Dan Kesehatan Fakultas Kedokteran Universitas Islam Sumatera Utara*, 24(1), 182–190.
- Wijayati, R., & Saputro, D. R. S. (2023). *Clustering* Data Campuran Numerik Dan Kategorik Menggunakan Algoritme *K-Prototype*. *PRISMA, Prosiding Seminar Nasional Matematika*, 6, 702–706.
- Yulnefia, & Sutia, M. (2022). Hubungan Riwayat Penyakit Infeksi Dengan Kejadian Stunting Pada Balita Usia 24-36 Bulan Di Wilayah Kerja Puskesmas Tambang Kabupaten Kampar. *JMJ*, 10(1).