JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Clustering Data Penyakit Pasien Pada Puskesmas Mulyaguna Menggunakan Algoritma K-Means

Muhammad Hafiz Ziqrullah[1], Andri[2*], Susan Dian Purnamasari[3],Ilman Zuhri Yadi[4]
[1,2,3,4]Sistem Informasi, Unversitas Bina Darma, Palembang, Indonesia
Email: [1]201410080@student.binadarma.ac.id, [2]andri@binadarma.ac.id,
[3]susandian@binadarma.ac.id,[4]ilmanzuhriyadi@binadarma.ac.id
Email Penulis Korespondensi: [1]201410080@student.binadarma.ac.id

Abstrak− Penelitian ini bertujuan untuk mengoptimalkan kegiatan penyuluhan di Puskesmas Mulyaguna dengan memanfaatkan metode clustering menggunakan algoritma K-Means, yang membagi data ke dalam kelompok berdasarkan kedekatan nilai atribut tertentu. Atribut yang digunakan dalam penelitian ini meliputi diagnosa, usia, dan jenis kelamin pasien. Hasil analisis menghasilkan enam klaster dengan kategori frekuensi tertinggi, tinggi, cukup tinggi, menengah, rendah, dan terendah. Klaster-klaster tersebut kemudian divisualisasikan untuk memberikan pemahaman yang lebih jelas mengenai distribusi data pasien, sehingga mempermudah penyusunan strategi penyuluhan yang lebih tepat sasaran di Puskesmas Mulyaguna.
Kata Kunci: Clustering, K-Means, Penyuluhan, Puskesmas Mulyaguna, Visualisasi

Abstract– This study aims to optimize counseling activities at the Mulyaguna Health Center by utilizing the clustering method using the K-Means algorithm, which divides data into groups based on the proximity of certain attribute values. The attributes used in this study include diagnosis, age, and gender of the patient. The results of the analysis produced six clusters with the highest, high, quite high, medium, low, and lowest frequency categories. These clusters were then visualized to provide a clearer understanding of the distribution of patient data, thus facilitating the preparation of more targeted counseling strategies at the Mulyaguna Health Center.
Keywords: Cluster, Counseling, K-Means, Mulyaguna Health Center, Visualization

1. PENDAHULUAN

Peraturan Menteri Kesehatan no 43 tahun 2019 berdasarkan pasal 1 menjelaskan bahwa "Puskesmas adalah fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif di wilayah kerjanya" [1]

Namun, berdasarkan wawancara dengan salah satu petugas di Puskesmas Mulyaguna, ditemukan bahwa upaya promotif dan preventif yang dilakukan selama ini terasa belum efektif. Informasi yang digunakan dalam penyuluhan hanya berupa data umum yang sederhana dan dalam jumlah besar, sehingga tidak memberikan dampak yang maksimal. Ketidakefektifan ini berpotensi mengurangi kualitas layanan kesehatan yang diberikan kepada masyarakat, disebabkan penyuluhan yang belum tentu tepat sasaran, tidak mampu memberikan informasi yang dibutuhkan oleh masyarakat. Akibatnya, sumber daya yang dimiliki oleh Puskesmas tidak termanfaatkan secara optimal, dan upaya pencegahan penyakit menjadi kurang efektif.

Untuk mengoptimalkan penyuluhan agar lebih tepat sasaran, puskesmas perlu mendapatkan informasi yang relevan melalui analisis data penyakit pasien yang berobat di Puskesmas sebelumnya. Sebagai contoh, ketika Puskesmas Mulyaguna merencanakan program penyuluhan tentang pencegahan penyakit, mereka bisa melakukan analisis data dengan memahami pola penyakit berdasarkan atribut diagnosa, usia, dan jenis kelamin. Dengan demikian, puskesmas dapat merancang penyuluhan yang lebih terarah dan efektif.

Pendekatan sistematis dalam menganalisis data diperlukan untuk menelusuri informasi yang tersirat pada data penyakit pasien. Pendekatan ini menjadi landasan bagi perumusan strategi yang lebih efektif. Metode yang dinilai relevan untuk kegiatan ini adalah Data Mining.

Data Mining adalah proses ekstraksi atau penambangan data pada suatu informasi yang besar dan belum diketahui sebelumnya, namun dapat dimengerti serta berguna dalam menghasilkan suatu keputusan bisnis [2]. Salah satu teknik dalam data mining yang bisa digunakan dalam penelitian ini ialah clustering sebagai metode untuk mengelompokkan data berdasarkan kesamaan pola atau atribut tertentu, mampu untuk memberikan pemahaman terhadap fenomena penyebaran penyakit.

Clustering bertujuan untuk pengelompokkan data yang memiliki karakteristik sama ke dalam satu kelompok atau tempat yang sama dan memisahkan data yang memiliki perbedaan karakteristik. Data akan dikelompokkan dalam satu kelompok atau cluster mempunyai tingkat kemiripan yang maksimum dan minimum [2]. Adapun pada teknik clustering dibutuhkan suatu langkah-langkah sistematis atau yang sering disebut algoritma dengan tujuan memecahkan permasalahan dan memastikan semua proses pengklusteran berjalan dengan semestinya.

Algoritma K-Means adalah metode klasterisasi non-hierarkis berbasis jarak yang membagi data ke dalam kelompok (cluster) dan beroperasi pada atribut numerik. Algoritma K-Means termasuk dalam klasterisasi partisi yang membagi data ke dalam area yang terpisah. Algoritma K-Means sangat populer karena kemudahannya dan kemampuannya untuk mengklaster data besar serta outlier dengan sangat cepat [3]

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Dengan definisi-defisini tersebut, dapat disimpulkan bahwa data mining mampu menjawab permasalahan yang ada pada proses strategi penyuluhan puskesmas agar lebih tepat sasaran. Dengan mengidentifikasi data pasien yang memiliki jumlah data yang besar, data mining memberikan solusi dengan mengelompokkan data-data tersebut dan menghasilkan informasi yang relevan untuk digunakan dalam merancang strategi penyuluhan.

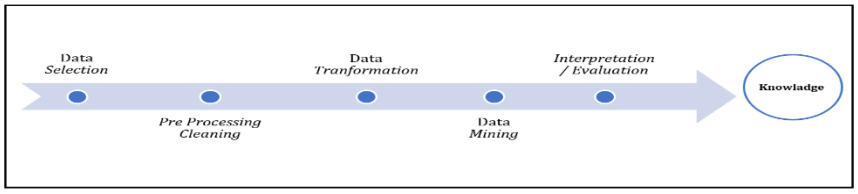
Misalnya, jika data menunjukkan bahwa penyakit ispa sering dialami oleh usia anak-anak dan lansia dengan jenis kelamin domininan yaitu laki-laki, mereka dapat fokus pada penyuluhan terkait penyakit ispa dan menargetkan peserta penyuluhan yang berada pada usia dan jenis kelamin pasien-pasien tersebut.

Pada penelitian sebelumnya yang berjudul "Pengelompokkan Penyakit Berdasarkan Lingkungan Dengan Algoritma K-Means Pada Puskesmas Sungai Tarab 2" [4]. penelitian tersebut berhasil membentuk cluster sesuai tujuan penelitian. Namun, hasil pengelompokan (clustering) yang diperoleh, disajikan dalam format yang tidak mudah dipahami. Hal ini menyebabkan pengguna membutuhkan waktu lebih lama untuk memahami pola data yang ada.

Untuk mengatasi kelemahan tersebut, penelitian ini memanfaatkan teknik visualisasi data sebagai solusi. Visualisasi data memungkinkan hasil pengelompokan data ditampilkan dalam bentuk grafik atau bagan, sehingga informasi lebih mudah dipahami dan digunakan[5]. Salah satu aplikasi atau perangkat lunak yang dapat mendukung berbagai macam proses pada visualisasi data ialah tableau, Tableau merupakan sebuah software yang memiliki berbagai macam alat visualisasi seperti grafik, diagram, pemetaan geografis dan sebagianya, sehingga dapat lebih mudah dipahami [6].

2. METODOLOGI PENELITIAN

Knowladge Discovery in Database merupakan metode analisa data dalam penelitian ini. Tahapan-tapahan analisa data dapat dilihat sebagai berikut :



Gambar 1. Tahapan Knowladge Discovert in Database (KDD)

2.1 Data Selection

Tahap ini mencakup pemilihan data yang akan digunakan dalam proses data mining, disesuaikan dengan informasi atau tujuan yang hendak dicapai [7]. Data yang digunakan adalah data pasien umum dan BPJS Puskesmas Mulyaguna tahun 2023 dengan 1789 record dan 15 atribut. Tahapan ini menyeleksi 3 atribut utama: diagnosa penyakit, usia, dan jenis kelamin.

2.2 Pre-Processing Cleaning

Pembersihan data melibatkan operasi dasar, seperti penghapusan noise dari data [8]. Tahapan ini berhasil membersihkan data dari 1778 menjadi 1769 record, dengan menghapus 8 data yang memiliki nilai null

2.3 Data Tranformation

Tahap ketiga adalah transformasi data agar sesuai dengan algoritma. Inisialisasi data karakteristik menjadi numerik penting dalam K-Means, karena algoritma ini menggunakan jarak Euclidean untuk mengukur kedekatan objek dengan centroid tiap cluster [9]. Jarak Euclidean hanya bisa dihitung pada data numerik, sehingga semua data atribut harus ditransformasi terlebih dahulu.

a. Tranformasi diagnosa (A1)

Tabel 1. Tranformasi Diagnosa

Diagnosa	Tranformasi
Ra (Radang Sendi)	1
D. Alergi	2
Skizopernia	3
Ispa	4
Hipertensi	5
Gastitis	6
Dermatitis	7

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Abses	8
Anemia	9
Chepalgia	10
Dispepsia	11
Faringitis	12
Febris	13
Ge	14
Kolestrol	15
Mialgia	16
Vertigo	17
Asam Urat	18
Suspect TB	19
Scables	20

b. Tranformasi Usia (A2)

Tabel 2. Tranformasi Usia [10]

Usia	Kelompok Usia	Tranformasi
1 – 14 Tahun	Anak-Anak	1
15 s/d 19	Remaja	2
20 s/d 44	Dewasa	3
>45	Lansia	4

c. Tranformasi Jenis Kelamin (A3)

Tabel 3. Tranformasi Jenis Kelamin

Diagnosa	Tranformasi
Laki-Laki	1
Perempuan	2

2.4 Data Mining

Pada tahap ini, proses data mining dijalankan dengan teknik Clustering dan algoritma K-Means menjadi bagian penting dari langkah-langkah dalam proses data mining.

Tahapan algoritma k-means [11], sebagai berikut ;

a) Menentukan jumlah cluster
b) Pilih titik pusat cluster (centroid) awal secara random
c) Hitung jarak data pada setiap titik pusat cluster (centroid)
d) Mengelompokkan data pada cluster terdekat
e) Tentukan centroid baru dengan menghitung rata-rata data di tiap cluster.
f) Hitung jarak data ke centroid baru dan kelompokkan kembali data ke cluster terdekat
g) Hentikan jika tidak ada perubahan anggota cluster, jika ada ulangi dari langkah 5.

2.5 **Interpretation / Evaluation**

Tahap ini dilakukan penyederhanaan informasi yang dihasilkan dari proses data mining harus disajikan dalam format yang mudah dipahami oleh pihak terkait [12].

2.6 Knowladge

Knowladge merupakan tahap terakhir dari metode Knowladge Discovery in Database, tahapan ini adalah menghasilkan informasi yang didapatkan dari proses data mining sebelumnya

3. HASIL DAN PEMBAHASAN

Mencakup 3 proses selanjutnya, yaitu Data Mining, Interpretation, dan Knowladge

3.1 Data Mining

Pada proses data mining menggunakan algoritma k-means untuk mengetahui cluster penyakit dengan frekuensi tertinggi hingga terendah.

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

3.1.1 Perhitungan Manual

Dilakukan uji coba dengan menggunakan 15 sampel data pasien yang diambil secara acak, data yang digunakan merupakan data yang telah ditranformasi menjadi numerik. Perlu diingat tahap uji coba dilakukan hanya untuk melihat proses perhitungan, bukan hasil sebenarnya.

Tabel 4. Sampel Data

Data	Diagnosa(A1)	Usia (A2)	Jenis Kelamin(A3)
1	1	4	2
2	2	3	2
3	2	4	2
4	4	4	2
5	4	4	2
6	2	1	2
7	4	4	2
21	4	3	2
369	6	4	1
1014	4	4	1
1058	8	3	2
1204	6	3	2
1243	6	3	1
1468	7	1	1
1533	4	1	2

a. Menentukkan jumlah cluster

Tentukkan jumlah cluster yang diinginkan, pada tahap uji coba menggunakan 4 cluter

b. Menetukan centorid awal

Tabel 5. Centroid Awal

Centroid	No	A1	A2	A3
Cluster_1	5	4	4	2
Cluster_2	21	4	3	2
Cluster_3	1014	4	4	1
Cluster_4	1533	4	1	2

c. Hitung jarak

Menghitung jarak data ke titik pusat cluster, dilakukan dengan menngunakan rumus euclidean distance. Berdasarkan penelitian [13]. Euclidean Distance lebih efektif dalam mengidentifikasi pola data dibanding Manhattan Distance dan Minkowski Distance.

Rumus Euclidience Distance :

$$D(i.j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \ldots \quad (1)$$

Ket : i = data, j = jarak

Perhitungan jarak terdekat data 1 ke titik pusat cluster :

d (1, 1) $= \sqrt{(1-4)^2 + (4-4)^2 + (2-2)^2} = 3$

d (1, 2) $= \sqrt{(1-4)^2 + (4-3)^2 + (2-2)^2} = 3{,}16227766017$

d (1, 3) $= \sqrt{(1-4)^2 + (4-4)^2 + (2-1)^2} = 3{,}16227766017$

d (1, 4) $= \sqrt{(1-4)^2 + (4-1)^2 + (2-2)^2} = 4{,}24264068712$

Perhitungan jarak terdekat data 2 ke titik pusat cluster :

d (2, 1) $= \sqrt{(2-4)^2 + (3-4)^2 + (2-2)^2} = 2{,}2360679775$

d (2, 2) $= \sqrt{(2-4)^2 + (3-3)^2 + (2-2)^2} = 2$

d (2, 3) $= \sqrt{(2-4)^2 + (3-4)^2 + (2-1)^2} = 2{,}44948974278$

d (2, 4) $= \sqrt{(2-4)^2 + (3-1)^2 + (2-2)^2} = 3{,}82842712475$

Perhitungan jarak terdekat data 21 ke titik pusat cluster :

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

d (21, 1) $= \sqrt{(4-4)^2 + (3-4)^2 + (2-2)^2} = 1$
d (21, 2) $= \sqrt{(4-4)^2 + (3-3)^2 + (2-2)^2} = 0$
d (21, 3) $= \sqrt{(4-4)^2 + (3-4)^2 + (2-1)^2} = 1{,}41421356237$
d (21, 4) $= \sqrt{(4-4)^2 + (3-1)^2 + (2-2)^2} = 2$

hitung semua data dengan jarak terdekat pada centroid sampai selesai sesuai sampel data yang diambil.

d. Mengelompokkan data pada cluster terdekat

Jarak yang paling dekat suatu data terhadap titik centroid setiap cluster akan menentukan posisi data tersebut untuk masuk ke cluster tertentu. Contohnya pada data ke_1 dengan jarak terdekat bernilai = 3, maka pasien tersebut masuk kedalam cluster 1. Berbeda dengan data ke-2 yang memiliki jarak terdekat bernilai = 2 maka pasien tersebut masuk kedalam cluster 2.

Tabel 6. Iterasi 1

Data	Jarak cluster 1	Jarak cluster 2	Jarak cluster 3	Jarak cluster 4	Hasil cluster
1	3	3,16227766017	3,16227766017	4,24264068712	C1
2	2,2360679775	2	2,44948974278	3,82842712475	C2
3	2	2,2360679775	2,2360679775	3,605551275465	C1
4	0	1	1	3	C1
5	0	1	1	3	C1
6	3,60555127546	2,82842712475	3,74165738677	2	C4
7	0	1	1	3	C1
21	1	0	1,41421356237	2	C2
369	2,2360679775	2,44948974278	2	3,74165738677	C3
1014	1	1,41421356237	0	3,16227766017	C3
1058	4,12310562562	4	4,24264068712	4,472135955	C2
1204	2,2360679775	2	2,449489742278	2,82842712475	C2
1243	2,44948974278	2	2,2360679775	3	C2
1468	4,35889894354	3,74165738677	4,24264068712	3,16227766017	C4
1533	3,16227766017	2,2360679775	3	0	C4

e. Menentukan centroid baru dilakukan dengan menghitung rata-rata dari setiap anggota cluster yang diambil dari setiap data atribut yang digunakan.

C1 = (3 ; 4 ; 2)
A1 = $\left(\frac{1+2+4+4+4}{5}\right) = 3$
A2 = $\left(\frac{4+4+4+4+4}{5}\right) = 4$
A3 = $\left(\frac{2+2+2+2+2}{5}\right) = 2$

C3 = (5 ; 4 ; 1)
A1 = $\left(\frac{6+4}{2}\right) = 5$
A2 = $\left(\frac{4+4}{2}\right) = 4$
A3 = $\left(\frac{1+1}{2}\right) = 1$

C2 = (5,2 ; 3 ; 1,8)
A1 = $\left(\frac{2+4+8+6+6}{5}\right) = 5{,}2$
A2 = $\left(\frac{3+3+3+3+3}{5}\right) = 3$
A3 = $\left(\frac{2+2+2+2+1}{5}\right) = 1{,}8$

C4 = (4,3 ; 2 ; 1,6)
A1 = $\left(\frac{2+7+4}{3}\right) = 4{,}3$
A2 = $\left(\frac{1+1+1}{3}\right) = 1$
A3 = $\left(\frac{2+1+2}{2}\right) = 1{,}6$

f. Menghitung dan Mengelompokkan kembali data kedalam cluster terdekat menggunakan centroid baru

Dilakukan perhitungan kembali jarak setiap data ke centroid baru dan menggelompokkan data yang memiliki jarak terdekat ke setiap cluster dengan cara yang sama dengan tahapan 3 dan 4

Tabel 7. Iterasi 2

Data	Jarak cluster 1	Jarak cluster 2	Jarak cluster 3	Jarak cluster 4	Hasil cluster
1	2	4,32203655699	3,80131556175	4,63680924775	C1
2	3,2	1,41421356237	2,69258240357	3,23264597505	C2
3	1	3,358571112475	2,87228132327	3,93700393701	C1
4	1	1,5748015748	1,3601705087	3,08220700148	C1
5	1	1,5748015748	1,3601705087	3,08220700148	C1

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

6	3,16227766017	3,77888872554	3,3510196625	2,5495097568	C4
7	1	1,5748015748	1,3601705087	3,08220700148	C1
21	1,41421356237	0,8602325267	1,21655250606	2,1432034356	C2
369	3,16227766017	1,,74642491966	1	3,31662479036	C3
1014	1,41421356237	1,75499287748	1	3,08220700148	C3
1058	5,09901951359	3,44093010682	3,4711099154	4,06201920232	C2
1204	3,16227766017	0,82462112512	1,5	2,5495097568	C2
1243	3,31662479036	1,1313708499	1,43178210633	2,5495097568	C2
1468	5,09901951359	2,80713376952	3,1384709653	2,5495097568	C4
1533	3,16227766017	2,46576560119	2,29346898824	0,70710678119	C4

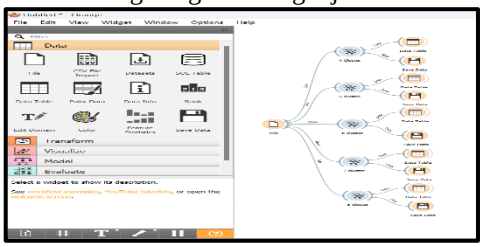
g. Perhitungan dihentikan karena tidak ada perubahan pada anggota cluster

3.1.2 Orange Data Mining

Orange adalah tools berbasis Python yang dapar digunakan untuk melakukan proses data mining [14]. Proses ini menggunakan record data sebanyak 1769 pasien dengan 3 atribut yaitu diagnosa, usia, dan jenis kelamin yang sudah dilakukan proses tranformasi menjadi data numerik.

a. Proses Orange Data Mining dan Menentukkann Cluster terbaik

Clustering data pasien pada orange data mining dengan berbagai jumlah cluster yaitu 4,5,6,7,dan 8.



Gambar 2. Orange Data Mining

Setiap hasil cluster akan dilakukan proses DBI untuk melihat kualitas jumlah cluster terbaik. Proses DBI dilakukan menggunakan software tambahan yaitu google collab

Tabel 8. Devise Bouldin Index

Jumlah cluster	Devise Bouldin Index
4	0.809408586
5	0.860090412
6	0.792629259
7	0.817770435

Pada devise boulding index melihat nilai terendah sebagai nilai terbaik [15]. Dari perhiungan nilai DBI dihasilkan jumlah *cluster* optimal sebanyak 6 dengan nilai 0,792629259

b. Hasil cluster

6 cluster merupakan jumlah cluster terbaik, hasil dari clustering tersebut adalah cluster_1 memiliki 161 item, cluster_2 memiliki 446 item, cluster_3 memiliki 154 item, cluster_4 memiliki 457 item, cluster_5 memiliki 287 item, dan cluster_6 memiliki 264 item.

3.2 Interpretation

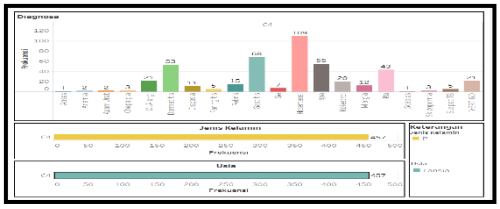
Pada tahapan ini ialah menampilkan hasil clustering ke bentuk visualisasi data, menggunakan software tableau dekstop. Untuk dapat menampilkan informasi yang mudah dipahami pengguna.

1. Visualisasi Frekuensi Pasien Tertinggi

Cluster_4 merupakan frekuensi pasien tertinggi memiliki 457 pasien berusia lansia dengan jenis kelamin perempuan.

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

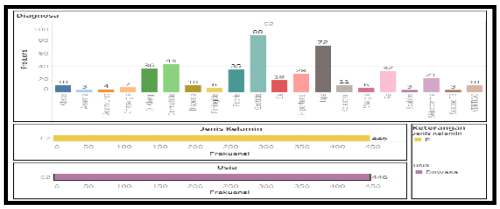
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 3. Visualisasi Frekuensi Pasien Teringgi

2. Visuaslisasi Frekuensi Pasien Tinggi

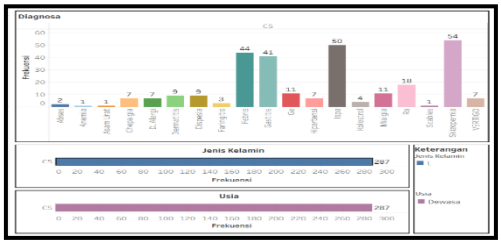
Cluster_2 merupakan frekuensi pasien tinggi memiliki 446 pasien berusia dewasa dengan jenis kelamin perempuan.



Gambar 4. Visualisasi Frekuensi Pasien Tinggi

3. *Visualisasi* Frekuensi Pasien Cukup Tinggi

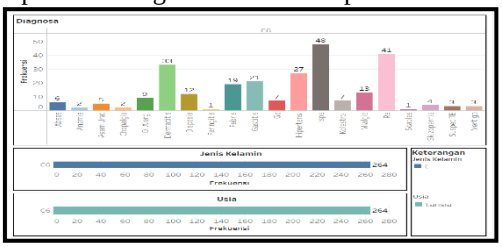
Cluster_5 merupakan frekuensi pasien cukup tinggi memiliki 287 pasien berusia dewasa dengan jenis kelamin laki-laki.



Gambar 5. Visualisasi Frekuensi Pasien Cukup Tinggi

4. Visualisasi frekuensi Pasien Menengah

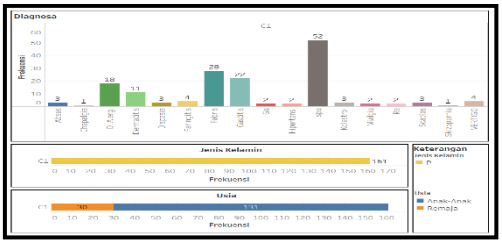
Cluster_6 merupakan frekuensi pasien menegah memiliki 264 pasien berusia lansia dengan jenis kelamin laki-laki.



Gambar 6. Visualisasi Frekuensi Pasien Menengah

5. Visualisasi Frekuensi Pasien Rendah

Cluster_1 merupakan frekuensi pasien rendah memiliki 161 pasien berusia anak-anak dan remaja dengan jenis kelamin perempuan.



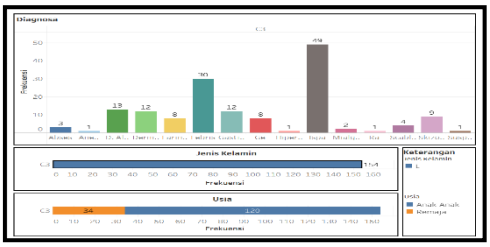
Gambar 7. Visualisasi Frekuensi Pasien Rendah

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 6 No. 2, 2024, Page 608-616
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

6. Visualisasi Frekuensi Pasien Terendah

Cluster_3 merupakan penyakit dengan frekuensi pasien terendah memiliki 154 pasien berusia anak-anak dan remaja dengan jenis kelamin laki-laki.



Gambar 8. Visualisasi Frekuensi Pasien Terendah

3.3 Knowladge

Berdasarkan penelitian yang telah diselesaikan, didapatkan sebuah pola dan knowladge baru dari proses data mining yang menggunakan algoritma k-means dalam pengelompokkan data pasien puskesmas mulyaguna pada tahun 2023, penelitian ini menghasilkan 6 cluster yaitu pasien dengan frekuensi penderita penyakit tertinggi, tinggi, cukup tinggi, menegah, rendah dan terendah.

1. Karakteristik setiap cluster

Tabel 9. Karakteristik Setiap Cluster

Frekuensi	Cluster	Diagnosa Terbanyak Setiap Cluster	Usia	Jenis Kelamin
Tertinggi	Cluster_4	Hipertensi, Gastitis, Ispa, Dermatitis, Ra	Lansia	Perempuan
Tinggi	Cluster_2	Gastitis, Ispa, Dermatitis, D Alergi, Ra	Dewasa	Perempuan
Cukup Tinggi	Cluster_5	Skizopernia, Ispa, Febris, Gastitis, Ra	Dewasa	Laki-Laki
Menengah	Cluster_6	Ispa, Febris, Gastitis, D Alergi, Dermatitis	Lansia	Laki-Laki
Rendah	Cluster_1	Ispa, Febris, Gastitis, D Alergi, Dermatitis	Anak-Anak dan Remaja	Perempuan
Terendah	Cluster_3	Ispa, Febris, D Alergi, Gastitis, Dermatitis	Anak-Anak dan Remaja	Laki-Laki

2. Diagnosa penyakit terbanyak keseluruhan

Tabel 10. Diagnosa Penyakit Pasien Terbanyak Keseluruhan

Diagnosa	Frekuensi
Ispa	326
Gastitis	252
Hipertensi	174
Dermatisis	161
Ra	137

KESIMPULAN

Penelitian ini berhasil mengelompokkan data penyakit pasien di Puskesmas Mulyaguna menjadi 6 cluster yaitu cluster dengan frekuensi pasien tertinggi, tinggi, cukup tinggi, menengah, rendah, dan terendah dengan menggunakan algoritma K-Means. Enam cluster tersebut mendapat skor Devise Bouldin Index terbaik dengan nilai 0.792629259.

Cluster dengan frekuensi pasien tertinggi menandakan bahwa kelompok ini memiliki jumlah pasien terbesar dengan karakteristik tertentu berdasarkan diagnosa, usia, dan jenis kelamin. Sebaliknya, cluster dengan frekuensi pasien rendah menunjukkan kelompok dengan jumlah pasien paling sedikit. Pengelompokan ini dapat digunakan sebagai dasar untuk perencanaan dan pengambilan keputusan yang lebih tepat terkait penanganan kesehatan di Puskesmas Mulyaguna.

Selain itu, penelitian ini juga mencakup visualisasi hasil clustering untuk mempermudah pemahaman pihak Puskesmas dalam membaca data cluster. Visualisasi ini dilakukan dengan menggunakan diagram yang menyajikan informasi tentang distribusi pasien dalam setiap cluster dengan cara yang jelas dan mudah dipahami.

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**

Dengan memahami distribusi pasien dalam setiap cluster, Puskesmas dapat lebih efektif dalam mengalokasikan sumber daya medis, melakukan pencegahan penyakit, dan merancang program kesehatan yang lebih sesuai dengan kebutuhan populasi pasien yang dilayani

UCAPAN TERIMAKASIH

Terimakasih kepada Universitas Bina Darma Palembang, Dosen Pembimbing yang telah mendukung dan membimbing dalam penelitian ini. Serta orangtua dan keluarga yang selalu memberikan semangat dan do'a sampai menyelesaikan penelitian ini.

REFERENCES

[1] Menteri Kesehatan RI, "Peraturan Menteri Kesehatan RI No 49 Tahun 2016 Tentang Pusat Kesehatan Masyarakat," Jakarta, 2016. [Daring]. Tersedia pada: www.peraturan.go.id

[2] N. L. W. S. R. Ginantara *dkk.*, *FullBook Data mining dan Penerapan Algoritma*, 1 ed. Yayasan Kita Menulis, 2021.

[3] B. Serasi Ginting dan M. Simanjuntak, "Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus : Puskesmas Bahorok)," *ALGORITMA: Jurnal Ilmu Komputer dan Informatika*, hlm. 2, 2021.

[4] D. P. Sari, "PENGELOMPOKKAN PENYAKIT BERDASARKAN LINGKUNGAN DENGAN ALGORITMA K-MEANS PADA PUSKESMAS SUNGAI TARAB 2," *JOISIE Journal Of Information System And Informatics Engineering*, vol. 5, no. Desember, hlm. 75–81, 2021.

[5] D. A. N. Rizki, "Visualisasi Data Sentimen Terhadap Organisasi Perangkat Daerah Pemerintah Provinsi Jawa Barat Di Jabar Digital Service," 2020, Diakses: 1 April 2024. [Daring]. Tersedia pada: https://elibrary.unikom.ac.id/id/eprint/2658/

[6] D. Saepuloh, "Visualisasi Data Covid 19 Provinsi DKI Menggunakan Tableau," *Jurnal Riset Jakarta*, vol. 13, no. 2, Des 2020, doi: 10.37439/jurnaldrd.v13i2.37.

[7] A. Srirahayu dan L. Setya Pribadie, "JURNAL ILMIAH INFORMATIKA GLOBAL Review Paper Data Mining Klasifikasi Data Mining", doi: 10.36982/jiig.v13i2.2307.

[8] "BAB II LANDASAN-TEORI 2.1 Dataset NSL-KDD." [Daring]. Tersedia pada: https://repository.uin-suska.ac.id/16170/7/7.%20BAB%20II.pdf

[9] N. Mara dan N. S. Intisari, "PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS," 2013.

[10] M. B. Fajri dan S. D. Purnamasari, "Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering," 2022. [Daring]. Tersedia pada: https://journal-computing.org/index.php/journal-ita/index

[11] M. Rachman Mulyandi *dkk.*, "Implementasi Algoritma K-Means Clustering Dalam," 2023.

[12] A. Srirahayu dan L. Setya Pribadie, "JURNAL ILMIAH INFORMATIKA GLOBAL Review Paper Data Mining Klasifikasi Data Mining", doi: 10.36982/jiig.v13i2.2307.

[13] B. Hartono, S. Eniyati, dan K. Hadiono, "Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 3, hlm. 503, Mar 2023, doi: 10.30865/json.v4i3.6021.

[14] M. Muharrom, "Analisis Penggunaan Orange Data Mining untuk Prediksi Harga USDT/BIDR Binance," *Bulletin of Information Technology (BIT)*, vol. 4, no. 2, hlm. 178–184, 2023, doi: 10.47065/bit.v3i1.

[15] S. Butsianto dan N. Saepudin, "PENERAPAN DATA MINING TERHADAP MINAT SISWA DALAM MATA PELAJARAN MATEMATIKA DENGAN METODE K-MEANS," 2019.

Submitted: **18/09/2024**; Accepted: **26/11/2024**; Published: **15/12/2024**