



Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan

Okta Jaya Harmaja¹, Hadirat Halawa², Wicarda Sandi Hulu³, Seriani Loi⁴

^{1,2,3} Sistem Informasi, Universitas Prima Indonesia, Jl. Sampul No. 4 Medan, Indonesia.

Email Korespondensi : [!Oktajaya.h@gmail.com](mailto:Oktajaya.h@gmail.com)

Abstract— Dengan meningkatnya kesadaran masyarakat tentang pentingnya perawatan kesehatan dan aksesibilitas yang lebih baik pada layanan kesehatan, jumlah pasien yang datang ke puskesmas dapat meningkat secara signifikan. Oleh karena itu, perlu solusi yang efisien untuk mengelompokkan jenis penyakit pasien agar pelayanan kesehatan dapat disesuaikan dengan kebutuhan yang terus berkembang. Implementasi *Algoritma K-Means Clustering* merupakan solusi efektif untuk mengatasi tantangan dalam mengelompokkan jenis penyakit pasien pada Puskesmas Pulo Brayan. Algoritma ini memungkinkan pengelompokan otomatis berdasarkan atribut yang relevan, meningkatkan efisiensi dalam pengelompokan pasien, dan memberikan wawasan yang lebih baik tentang kondisi kesehatan masyarakat setempat. Pada analisis ini, pengelompokan penyakit pasien dilakukan berdasarkan usia, jenis kelamin dan diagnosis penyakit. Penelitian ini menggunakan aplikasi RapidMiner 10.1.2. dengan menghasilkan 3 cluster dari 949 data pasien. cluster rendah sebanyak 198 pasien, cluster sedang sebanyak 227 pasien, dan cluster tinggi sebanyak 524 pasien. Cluster tinggi merupakan data yang paling sering dialami oleh pasien dengan jumlah 524 pasien, Cluster ini di dominasi oleh pasien laki-laki dengan rentang usia 40 tahun keatas, Diagnosa yang paling umum dialami oleh pasien dalam cluster ini adalah penyakit ISPA (Infeksi Saluran Pernapasan Akut)

Kata Kunci: Puskesmas, K-Means Clustering, Data Mining, Pengelompokan Penyakit.

Abstract— With increasing public awareness about the importance of health care and better accessibility to health services, the number of patients coming to the puskesmas can increase significantly. Therefore, an efficient solution is needed to classify the type of patient's disease so that health services can be adapted to the needs that continue to grow. The implementation of the K-Means Clustering Algorithm is an effective solution to overcome challenges in classifying the types of patient diseases at the Pulo Brayan Health Center. This algorithm enables automatic grouping based on relevant attributes, improves efficiency in patient categorization, and provides better insight into local public health conditions. In this analysis, patient grouping was done based on age, sex and disease diagnosis. This study uses the RapidMiner 10.1.2 application. by producing 3 clusters of 949 patient data. the low cluster is 198 patients, the medium cluster is 227 patients, and the high cluster is 524 patients. The high cluster is the data most often experienced by patients with a total of 524 patients. This cluster is dominated by male patients with an age range of 40 years and over. The most common diagnosis experienced by patients in this cluster is ARI (Acute Respiratory Infection) disease.

Keywords: Public Health Center, K-Means Clustering, Data Mining, Disease Categorization.

I. PENDAHULUAN

Puskesmas merupakan unit organisasi pelayanan kesehatan terdepan dengan misi sebagai pusat pengembangan pelayanan kesehatan, yang tugasnya melaksanakan pembinaan, pelayanan kesehatan secara menyeluruh dan terpadu kepada masyarakat di suatu wilayah tertentu[1]. Puskesmas Pulo Brayan merupakan salah satu unit pelayanan kesehatan masyarakat yang setiap harinya melayani banyak pasien dengan berbagai jenis penyakit sehingga pihak Puskesmas Pulo Brayan dapat menghasilkan data kunjungan pasien yang cukup banyak dan semakin bertambah tiap tahunnya sehingga mengakibatkan Puskesmas Pulo Brayan menghadapi tantangan dalam mengelola jumlah pasien yang terus bertambah dari waktu ke waktu. Dengan meningkatnya kesadaran masyarakat tentang pentingnya perawatan kesehatan dan aksesibilitas yang lebih baik pada layanan kesehatan, jumlah pasien yang datang ke puskesmas dapat meningkat secara signifikan. Oleh karena itu perlu solusi yang efisien untuk mengelompokkan jenis penyakit pasien agar pelayanan kesehatan dapat disesuaikan dengan kebutuhan yang terus berkembang. Selain itu Puskesmas Pulo Brayan melayani pasien dengan berbagai jenis penyakit yang bervariasi. Keberagaman ini membuat

sulit untuk mengorganisir pasien dan mengidentifikasi pola atau tren yang mungkin ada di antara data kunjungan pasien. Penting untuk mengelompokkan penyakit secara efisien sehingga dapat diperoleh wawasan yang berguna dalam pengambilan keputusan kesehatan. efisiensi dan produktivitas pelayanan kesehatan merupakan aspek penting yang perlu diperhatikan. Untuk meningkatkan efisiensi dan produktivitas, diperlukan solusi yang dapat mengotomatisasi proses ini. pemahaman yang lebih baik tentang kesehatan masyarakat menjadi faktor penting. Dengan mengelompokkan jenis penyakit pasien.

Implementasi *Algoritma K-Means Clustering* merupakan solusi efektif untuk mengatasi tantangan dalam mengelompokkan penyakit pasien pada Puskesmas Pulo Brayan. Algoritma ini memungkinkan pengelompokan otomatis berdasarkan atribut yang relevan, meningkatkan efisiensi dalam pengelompokan pasien, dan memberikan wawasan yang lebih baik tentang kondisi kesehatan masyarakat setempat. Beberapa penelitian sebelumnya telah banyak menggunakan Algoritma K-means, diantaranya Penelitian terdahulu pengelompokkan Algoritma K-Means [2]. Pada penelitian ini menggunakan implementasi algoritma k-means Clustering untuk mengetahui kelompok jenis penyakit



yang paling sering di alami oleh pasien. Penelitian lain menegompokkan penyakit[3]. Pada penelitian ini menggunakan implementasi algoritma K-Means Clustering dengan menggunakan 2 cluser yang mana pada cluster 0 mengelompokkan penyakit akut, sedangkan cluster 1 mengelompokkan penyakit tidak akut. Penerapan Algoritma k-means clustering untuk pengelompokkan penyebaran diare di kabupaten langkat[4] berdasarkan penelitian tersebut menghasilkan titik pusat cluster 1 (2.75 ; 3.75) atau pusat cluster berada pada kecamatan Batang serangan, Brandan Barat dan pematang jaya, sedangkan pusat cluster 2 : (2.93 ; 3.60) berada pada kecamatan Hinai, Sei Bingai dan Sirapit[4]. Selanjutnya algoritma k means clustering juga di terapkan pada penelitian dengan judul penerapan algoritma k-means clustering untuk pengelompokkan penyebaran covid-19 di provinsi jawa barat[5]. Pada penelitian lainnya juga menggunakan algoritma k-means clustering untuk klasifikasi penyakit pada daun stroberi menggunakan k-means clustering dan jaringan saraf tiruan [6].

Pada permasalahan diatas Penulis berupaya mengaplikasikan data mining sebagai solusi dalam mengelola data pasien yang bertumpuk serta jenis penyakit yang beragam. Dengan didukung oleh perangkat lunak yang kini dapat digunakan dengan mudah serta menjadikan efisiensi waktu yang baik. Data yang diolah adalah data penyakit pasien yang dikumpulkan selama 3 bulan terakhir. Dataset ini terdiri dari 949 data dengan atribut seperti nama, jenis kelamin, umur, alamat, dan diagnosa pasien. Clustering mengelompokkan menjadi beberapa kategori berdasarkan kesamaan atribut. Tujuan dari metode ini adalah untuk mengetahui perbedaan antar kategori [7] Pada metode Clustering data atau nilai yang ada belum memiliki target sehingga memerlukan prediksi ke kategori mana data itu akan masuk [8].

Penelitian yang penulis lakukan ini mempunyai tujuan yaitu implementasikan metode *Clustering* dengan *Algoritma K-Means* pada data penyakit untuk mengelompokkan penyakit sehingga pihak puskesmas dapat meningkatkan pelayanannya dan memberikan pemahaman yang lebih baik tentang variasi penyakit pasien.

II. METODE PENELITIAN

2.1 Jenis Penelitian

Jenis penelitian yang akan dilakukan oleh peneliti adalah termasuk ke dalam jenis penelitian kualitatif. penelitian kualitatif adalah penelitian yang bermaksud untuk memahami fenomena tentang apa yang dialami oleh subjek penelitian seperti perilaku, persepsi, motivasi, tindakan dan lain-lain secara holistik dan dengan cara deskripsi dalam bentuk kata-kata dan bahasa, pada suatu konteks khusus yang alamiah dengan memanfaatkan berbagai metode alamiah[9]. Pada umumnya penelitian kualitatif memperoleh data utama dari wawancara dan observasi. Setelah itu peneliti akan menganalisis data

yang diperoleh. Jika hasil penelitian bertentangan dengan teori yang digunakan, kemungkinan akan muncul konsep atau teori baru sebagai hasil dari penelitian tersebut. Dalam jenis penelitian ini dapat membantu memaksimalkan peneliti dalam implementasi *K-Means Clustering* untuk mengelompokkan penyakit pasien pada Puskesmas Pulo Brayan.

2.2 Teknik Pengumpulan Data

Proses pengumpulan data mempunyai peranan yang penting dalam penelitian ini, karena metode pengumpulan data akan menentukan kualitas dan akurasi data yang akan dikumpulkan selama proses penelitian dengan berbagai macam metode pengumpulan data, penulis akan menggunakan metode pengumpulan data sebagai berikut[10].

1. Observasi

Pengumpulan data dengan cara observasi oleh peneliti yaitu dengan mengunjungi Puskesmas Pulo Brayan secara langsung guna untuk mendapatkan informasi tentang topik penelitian yang terkait.

2. Wawancara

Peneliti melakukan wawancara kepada Ibu Eti sebagai bagian Tata Usaha pada Puskesmas Pulo Brayan guna meminta izin mengadakan penelitian dan memberitahukan tujuan dari pada penelitian tersebut serta memperoleh informasi tentang data penyakit pasien yang ada pada Puskesmas Pulo Brayan. Wawancara juga dilakukan peneliti kepada beberapa ibu yang merupakan tenaga medis pada Puskesmas Pulo Brayan guna memperoleh informasi terkait dengan penelitian yang dilakukan.

3. Studi Literatur

Untuk mendukung penelitian ini, peneliti melakukan pengumpulan referensi dari berbagai sumber, seperti buku, ebook, jurnal nasional dan internasional, serta berbagai sumber lainnya yang relevan dengan topik data mining, *Algoritma K-Means clustering* dan aplikasi RapidMiner

2.3 Prosedur Kerja

Penelitian ini mengadopsi metode Cross Industry Standard Process for Data Mining (CRISP-DM) untuk mengimplementasikan algoritma K-means clustering dalam pengelompokkan jenis penyakit pada Puskesmas Pulo Brayan. CRISP-DM adalah metode yang di gunakan penulis dalam penelitian ini. CRISP-DM merupakan suatu metodologi data mining yang disusun oleh konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam



data mining, Menurut Larose, data mining memiliki enam fase CRISP-DM, yaitu [11].

1. Business Understanding

Pada tahap ini penulis melakukan pemahaman atau analisa latar belakang pada Puskesmas Pulo Brayon guna untuk mengidentifikasi permasalahan yang ada, Kemudian digunakan sebagai acuan untuk menentukan tujuan dan perencanaan strategi dalam penerapan proses pemodelan data mining.

2. Data Understanding

Pada tahap ini penulis melakukan proses pengumpulan data awal pada puskesmas Pulo Brayon. kemudian melakukan pemahaman data yang bertujuan untuk mengenali data yang akan di terapkan dalam proses impementasi data mining. setelah melakukan tahap pemahaman data. data yang akan di terapkan penulis adalah data pasien penyakit pada puskesmas padang bulan.

3. Data Preparation

Data Preparation merupakan proses persiapan data yang meliputi pembersihan, integrasi, transformasi dan pemilihan atribut yang relevan untuk analisis lebih lanjut. Tujuannya adalah memastikan kualitas data yang baik dan siap digunakan dalam tahap modelling.

4. Modelling

Pada tahap ini melibatkan dan menerapkan beberapa teknik pemodelan dan sejumlah parameter yang disesuaikan sehingga dapat nilai yang optimal. Adapun pemodelan yang di lakukan oleh penelitian ini adalah model Clustering menggunakan Algoritma K-Means untuk mengelompokkan jenis penyakit pasien. Pada tahap ini model Clustering akan dibangun berdasarkan data yang telah dipersiapkan sebelumnya. Proses ini bertujuan untuk mengidentifikasi pola atau kelompok penyakit yang ada dalam dataset.

5. Evaluation

Dalam tahap ini fokus terhadap model yang didapatkan telah sesuai satandar K-Means Clustering yang telah dibangun dan tiada yang terlewat pada tahapan awal sampai tahap pemodelan selesai dan memenuhi tujuan yang telah ditetapkan sebelumnya.

6. Deployment

pada terakhir adalah implementasi hasil analisis dan penggunaan informasi yang diperoleh dari model Clustering. Informasi ini dapat digunakan oleh tim medis Puskesmas Pulo Brayon untuk mengambil keputusan kesehatan yang lebih baik, merencanakan program pencegahan dan pengobatan yang lebih terarah serta memberikan edukasi yang sesuai kepada pasien dan masyarakat setempat.

2.4 K-Means Clustering

K-Means adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981[12]. Clustering merupakan proses pengelompokkan objek ke dalam suatu cluster dimana objek yang berada dalam satu cluster mempunyai tingkat kesamaan yang tinggi dan antar cluster mempunyai tingkat kesamaan yang rendah[13]. Algoritma ini, mengklasterisasi berdasarkan tingkat kesesuaian dan ketidaksesuaian objek dengan cara membagi suatu objek menjadi *k-cluster* yang nantinya posisi objek tersebut dikelompokkan dari nilai mean terdekat[14]. Proses klasterisasi *K-Means* bergantung pada data yang ada serta kesimpulan yang ingin diambil Data yang sudah diatur di tiap-tiap pengamatan cluster berdasarkan ketetangaan pengamatan nilai mean kemudian diterapkan perhitungan berulang[15].

III. HASIL DAN PEMBAHASAN

3.1 Implementasi algoritma k-means

Pada penelitian ini, data yang akan digunakan untuk melakukan pemodelan atau implementasi Algoritma K-Means adalah data kunjungan pasien pada Puskesmas Pulo Brayon dari tanggal 1 Maret hingga 31 Mei 2023. Jumlah data yang tersedia sebanyak 949 data. kemudian data tersebut akan dikelompokkan kedalam 3 kelompok berdasarkan atribut jenis kelamin, umur dan diagnosa. Kemudian data tersebut diubah menjadi tipe numerik. Selanjutnya Algoritma K-Means akan digunakan untuk melakukan pengolahan data tersebut. Tabel 3.1 berisi data penyakit pasien pada puskesmas Pulo Brayon dapat dilihat sebagai berikut:

Tabel 1. Data penyakit Pasien Pada Puskesmas Pulo Brayon

No	Nama	Jenis Kelamin	Usia	Diagnosa
1	P1	P	8	Diare
2	P2	L	46	Ht
3	P3	L	35	Hordeolum Internum
4	P4	P	39	Ispa
5	P5	P	8	Ispa
6	P6	L	6	Ispa
7	P7	P	58	Dm + Ht
8	P8	L	38	Saijo
9	P9	L	53	Dm
10	P10	P	25	Dm
11	P11	P	30	Lbp
12	P12	P	42	Gi Poao
13	P13	P	36	Dm
14	P14	L	40	Prb
15	P15	P	59	Dm + Ht
16	P16	P	26	Dermatitis
17	P17	P	2	Diare
18	P18	L	51	Gout
19	P19	L	7	Tinea Corporis
20	P20	L	23	Tb



Berikut ini adalah data atribut yang akan diubah menjadi tipe numerik dan dilakukan pengkodean atau kategorikal nilai atribut pada data kunjungan puskesmas Pulo Brayan untuk pengelompokan penyakit pasien menggunakan Algoritma KMeans Clustering:

Tabel 2. Atribut Jenis Kelamin

Atribur jenis kelamin	
Laki – laki	1
Perempuan	2

Tabel 3. Atribut Usia (Kategorikal)

Atribut Usia	
0-10 tahun	1
11-25 tahun	2
26-40 tahun	3
>40 tahun	4

Tabel 4. Atribut Diagnosa Penyakit

Atribut Diagnosa Penyakit			
1	Ispa	11	Diare
2	Ispa + Dermatitis	12	Diare + Demam
3	Ispa+Diare	13	Dm + Ht
4	Abses	14	Dm
5	Artritis	15	Dispepsia
6	Cad	16	Dysesthesia
7	C Con	17	TB
8	Cesnound	18	Anemia
9	Cht	19	Epilepsi
10	Dermatitis	20	Febris

Dengan melakukan pengkodean dan pengkategorian ini, data atribut kunjungan pasien pada Puskesmas Pulo Brayan siap untuk digunakan dalam pengelompokan menggunakan Algoritma K-Means Clustering.

Tabel 5. Sampel Data Implemtasi K-Means Clustering

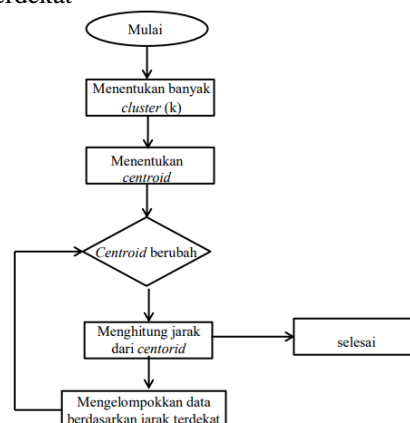
Nama	Jenis kelamin	Usia	Diagnosa
P1	2	1	11
P2	1	4	27
P3	1	3	28
P4	2	3	1
P5	2	1	1
P6	1	1	1
P7	2	4	13
P8	1	3	47
P9	1	4	14
P10	2	2	14
P11	2	3	43
P12	2	4	29

P13	2	3	14
P14	1	3	48
P15	2	4	13
P16	2	3	10
P17	2	1	11
P18	1	4	30
P19	1	1	53
P20	1	2	54

3.1 Perhitungan Algoritma K-Means Clustering

Gambar 3.1 dibawah ini menunjukkan diagram alur metode k-means yang digunakan dalam pengelompokan penyakit pada Puskesmas Pulo Brayan. Secara umum kinerja metode K-Means dapat dijelaskan dalam urutan langkah-langkah berikut[2]:

1. Menentukan banyaknya cluster (k)
2. Menentukan centroid (3)
3. Apakah nilai centroid berubah?
 - a. Jika ya, hitung jarak data dari centroid
 - b. Jika tidak, selesai.
4. Mengelompokkan data berdasarkan jarak terdekat



Gambar 1. Diagram Implementasi Algoritma K-Means Clustering

1. Penentuan pusat awal cluster

Pemilihan pusat awal cluster dilakukan secara acak pada data sampel implementasi Algoritma K-Means Clustering sehingga dapat memastikan objektivitas dan menghindari bias dalam pemilihan awal cluster. Dalam proses ini, tidak ada pertimbangan khusus terhadap atribut atau karakteristik data yang dipilih sebagai pusat awal cluster.

Tabel 6. Titik Pusat Awal Cluster

Centroid	Jenis Kelamin	Usia	Diagnosa
C1	1	3	28
C2	1	1	1
C3	2	2	14

Pusat awal cluster yang telah ditentukan akan digunakan sebagai titik awal dalam proses



pengelompokan data ke dalam cluster-cluster yang sesuai menggunakan metode K-Means.

2. Perhitungan jarak pusat cluster

Perhitungan jarak antara data dan pusat cluster dilakukan menggunakan metrik jarak yang tepat, dengan menggunakan persamaan *Euclidean Distance*. Metrik jarak ini berfungsi untuk mengukur seberapa dekat atau jauh suatu data dengan pusat cluster.

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan : X merupakan data dan y merupakan pusat cluster.

Jarak dihitung dengan memperhitungkan perbedaan nilai atribut antara data dan pusat cluster pada setiap dimensi atau atribut yang relevan.

Sehingga mendapatkan nilai matrik jarak sebagai berikut : jarak data ke-1 ke pusat cluster

$$C1 = \sqrt{(2 - 1)^2 + (1 - 3)^2 + (11 - 28)^2}$$

$$C1 = 17.1464282$$

$$C2 = \sqrt{(2 - 1)^2 + (1 - 1)^2 + (11 - 1)^2}$$

$$C2 = 10.04987562$$

$$C3 = \sqrt{(2 - 2)^2 + (1 - 2)^2 + (11 - 14)^2}$$

$$C3 = 3.16227766$$

Jarak data ke-2 ke pusat cluster

$$C1 = \sqrt{(1 - 2)^2 + (4 - 3)^2 + (27 - 28)^2}$$

$$C1 = 1.414213562$$

$$C2 = \sqrt{(1 - 1)^2 + (4 - 4)^2 + (27 - 1)^2}$$

$$C2 = 26.17250466$$

$$C3 = \sqrt{(1 - 1)^2 + (4 - 3)^2 + (27 - 14)^2}$$

$$C3 = 13.190906$$

Selanjutnya, perhitungan jarak dilanjutkan untuk setiap data ke-3 dan data seterusnya terhadap pusat awal cluster. Proses ini akan menghasilkan matriks jarak yang mencerminkan jarak antara setiap data dengan pusat awal cluster.

3. Pengelompokan data

Setelah dilakukan perhitungan jarak antara data dan pusat cluster, dilanjutkan dengan perbandingan untuk menentukan jarak terdekat antara data dengan pusat cluster. Jarak tersebut akan mengindikasikan bahwa data-data yang memiliki jarak terdekat akan tergabung dalam satu kelompok dengan pusat cluster yang terdekat, pengelompokan data tersebut dapat dilihat pada tabel sebagai berikut :

Tabel 7. Pengelompokan Jarak Terkecil Pusat Cluster Iterasi Ke-1

Nama	C1	C2	C3	Cluster
P1	17.1464282	10.04987562	3.16227766	3
P2	1.414213562	26.17250466	13.190906	1

P3	0	27.07397274	14.0712473	1
P4	27.01851217	2.236067977	13.0384048	2
P5	27.09243437	1	13.0384048	2
P6	27.07397274	0	13.0766968	2
P7	15.06651917	12.40967365	2.23606798	3
P8	19	46.04345773	33.0302891	1
P9	14.03566885	13.34166406	2.23606798	3
P10	14.07124728	13.07669683	0	3
P11	15.03329638	42.05948169	29.0172363	1
P12	1.732050808	28.17800561	15.132746	1
P13	14.03566885	13.19090596	1	3
P14	20	47.04253395	34.0293991	1
P15	15.06651917	12.40967365	2.23606798	3
P16	18.02775638	9.273618495	4.12310563	3
P17	17.1464282	10.04987562	3.16227766	3
P18	2.236067977	29.15475947	16.1554944	1
P19	25.07987241	52	39.0256326	1
P20	26.01922366	53.00943312	40.012498	1

Berdasarkan keterangan diatas maka dihasilkan kelompok sebagai berikut:

$$C1 = (p2, p3, p8, p11, p12, p14, p18, p19, p20)$$

$$C2 = (p4, p5, p6)$$

$$C3 = (p1, p7, p9, p10, p13, p15, p16, p17)$$

4. Penentuan pusat cluster baru

Setelah diperoleh anggota dari setiap cluster, langkah selanjutnya adalah menghitung pusat cluster baru berdasarkan data anggota dari setiap cluster yang telah ditemukan. Perhitungan ini menggunakan rumus yang sesuai dengan pusat anggota cluster, sebagai berikut:

$$C1 = \left(\begin{array}{c} \frac{1+1+1+2+2+1+1+1+1}{9}; \\ \frac{4+3+3+3+4+3+4+1+2}{8}; \\ \frac{27+28+47+43+29+48+30+54+54}{9}; \end{array} \right)$$

$$C1 = (0 ; 0 ; 39.88889)$$

$$C2 = \left(\frac{2+2+1}{3}; \frac{3+1+1}{3}; \frac{1+1+1}{3} \right)$$

$$C2 = (0 ; 0 ; 1)$$



$$C3 = \left(\begin{array}{c} \frac{2+2+1+2+2+2+2+2}{8}; \\ \frac{1+4+4+2+3+4+3+1}{8}; \\ \frac{11+13+14+14+14+13+10+11}{8} \end{array} \right)$$

$$C3 = (0 ; 0 ; 12,5)$$

Dari perhitungan yang telah dilakukan, diperoleh pusat *cluster* baru yang direpresentasikan dalam matriks pada tabel berikut:

Tabel 8. Pusat *Cluster* Baru

Centorid	Jenis Kelamin	Usia	Diagnosa
C1	0	0	39.88889
C2	0	0	1
C3	0	0	12.5

Iterasi berikutnya dilakukan dengan menggunakan langkah yang sama hingga tidak ada perubahan data yang terjadi dalam suatu *cluster*. perhitungan akan dihentikan jika tahap iterasi telah mencapai hasil yang sama, di mana tidak ada objek yang berpindah antar kelompok. Pada penelitian ini, proses iterasi dilakukan hingga mencapai iterasi ke-2. Hasil dari iterasi ke-2 dapat dilihat dalam tabel berikut:

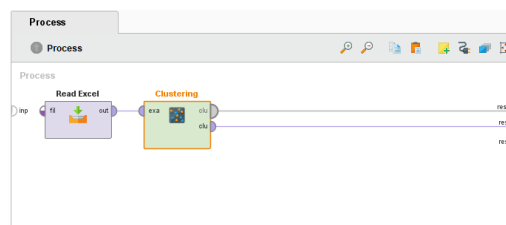
Tabel 9. Hasil Pengelompokkan Itarasi Ke-2

Nama	C1	C2	C3	Cluster
P1	28.97529812	10.24695077	2.692582404	3
P2	13.53231158	26.32489316	15.07481343	1
P3	12.30226317	27.18455444	15.81929202	1
P4	39.05567409	3.605551275	12.05197079	2
P5	38.95312156	2.236067977	11.71537451	2
P6	38.91459468	1.414213562	11.58663023	2
P7	27.2582528	12.80624847	4.5	3
P8	7.782538226	46.10856753	34.6446244	1
P9	26.21515912	13.6381817	4.387482194	3
P10	26.04293701	13.3041347	3.201562119	3
P11	4.762248665	42.15447782	30.71237536	1
P12	11.77148679	28.35489376	17.095321	1
P13	26.13875605	13.49073756	3.905124838	3
P14	8.70575232	47.10626285	35.64056677	1
P15	27.2582528	12.80624847	4.5	3
P16	30.10557555	9.695359715	4.387482194	3

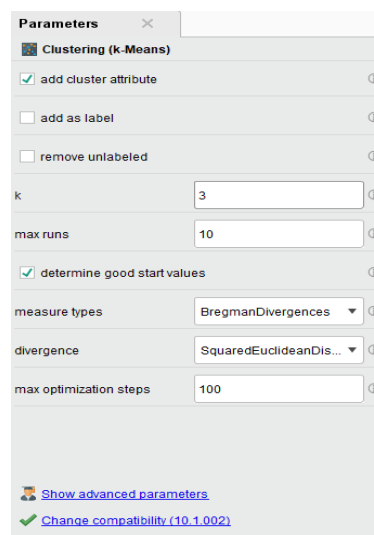
P17	28.97529812	10.24695077	2.692582404	3
P18	10.71401528	29.29163703	17.9791546	1
P19	13.18716173	52.01922721	40.52468384	1
P20	14.28717806	53.04714884	41.5601973	1

3.2 Implementasi *RapidMiner*

Pada proses awal melakukan modelling, mencari modeling, read *excel* lalu drag ke dalam proses[16]. Setelah itu, dilakukan import dataset penyakit pasien pada Puskesmas Pulo Brayon yang berformat *Excel* kedalam *RapidMiner*. Selanjutnya, tahap modeling menggunakan *Algoritma K-Means* dilakukan. Selanjutnya, data diproses menggunakan metode k-means dengan parameter k = 3 dan jenis pengukuran (measure type) campuran. Hal ini dilakukan karena data di Oj Cell terdiri dari kombinasi kode dan angka. Untuk menghitung jarak antara titik-titik cluster, digunakan Euclidean Distance sebagai standar.



Gambar 2. Pemodelan *K-Means Clustering* Menggunakan *RapidMiner*



Gambar 3. Parameter Pada *RapidMiner*

Selanjutnya, data diproses menggunakan *RapidMiner* dan menghasilkan tiga model klaster. Hasilnya adalah sebagai berikut: *cluster* 0 dengan 524 item, *cluster* 1 dengan 198 item, dan *cluster* 2 dengan 227 item. Perlu diperhatikan bahwa penomoran *cluster* dimulai dari 0, sesuai dengan konvensi dalam pemrograman di mana angka 0 merupakan angka pertama dalam urutan penomoran.



Cluster Model

Cluster 0: 524 items
 Cluster 1: 198 items
 Cluster 2: 227 items
 Total number of items: 949

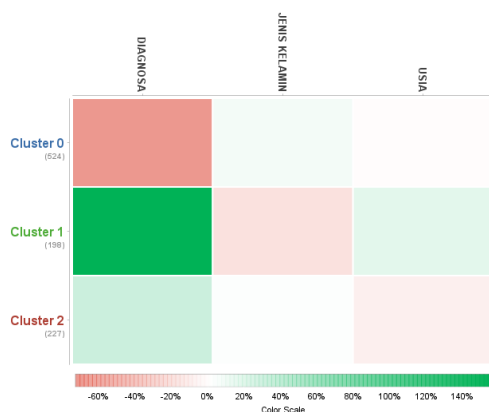
Gambar 4. Hasil Cluster Model

Attribute	cluster_0	cluster_1	cluster_2
JENIS KELAMIN	1.615	1.449	1.581
USIA	2.794	3.146	2.630
DIAGNOSA	6.055	48.207	25.251

Gambar 5. Hasil Centorid Table

3.3.1 Visualisasi Rapidminer

pada penelitian ini penulis melakukan proses visualisasi dengan menambahkan operator *cluster model visualizer* pada *RapidMiner* sehingga mendapatkan hasil sebagai berikut :



Gambar 6. Tampilan Visualisasi Heat Map

IV. KESIMPULAN

V. Kesimpulan

Hasil implementasi algoritma k-means clustering pada penelitian ini memberikan informasi penting dalam mengelompokkan data pasien menjadi tiga cluster berbeda, yaitu cluster rendah, cluster sedang, dan cluster tinggi. Masing-masing cluster mengandung 198 data pasien, 227 data pasien, dan 524 data pasien. Analisis lebih lanjut terhadap setiap cluster mengungkapkan bahwa cluster tinggi, atau cluster 1, merupakan kelompok yang paling sering dihadapi oleh pasien dengan total 524 pasien. Dominasi pasien laki-laki dengan rentang usia 40 tahun ke atas serta tingginya prevalensi penyakit ISPA (Infeksi Saluran Pernapasan Akut) menjadi ciri utama dalam cluster ini. Implementasi pengelompokkan

penyakit menggunakan algoritma K-Means dengan tiga kelompok di Puskesmas Pulo Brayon memiliki manfaat signifikan dalam memberikan pemahaman mendalam tentang variasi jenis penyakit pasien.

Temuan ini berpotensi menjadi landasan untuk merumuskan program pencegahan, pengobatan, dan edukasi yang lebih terarah untuk masing-masing kelompok pasien. Selain itu, penelitian ini dapat berkontribusi pada perumusan kebijakan kesehatan yang lebih tepat dan alokasi sumber daya yang lebih efektif, karena pemahaman yang lebih baik mengenai distribusi dan karakteristik penyakit akan memungkinkan peningkatan dalam pengelolaan fasilitas kesehatan dan penerapan strategi penanganan yang lebih efisien. Dengan demikian, hasil penelitian ini memiliki implikasi penting dalam meningkatkan kualitas pelayanan kesehatan dan upaya meningkatkan kesehatan masyarakat secara keseluruhan.

VI.REFERENSI

- [1] David Benyamin, "Pelayanan Kesehatan Masyarakat Di Puskesmas Sempaja Kecamatan Samarinda Utara," *Ejurnal Adm. Negara*, vol. 1, no. 2, pp. 440–452, 2013.
- [2] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," *Semin. Nas. Inov. Teknol.*, no. 2549–7952, pp. 285–291, 2020.
- [3] S. Maulia, B. Serasi Ginting, and A. Sihombing, "Implementasi Data Mining Pengelompokan Jenis Penyakit Pasien Menggunakan Metode Clustering (Studi Kasus : Puskesmas Sambirejo)," *J. Inform. Kaputama*, vol. 5, no. 1, pp. 71–80, 2021.
- [4] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, pp. 108–119, 2016, doi: 10.31154/cogito.v2i2.19.108-119.
- [5] N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.
- [6] *et al.*, "Klasifikasi Penyakit Pada Daun Stroberi Menggunakan K-Means Clustering dan Jaringan Syaraf Tiruan," *J. Keteknikan Pertan. Trop. dan Biosist.*, vol. 8, no. 2, pp. 161–170, 2020, doi: 10.21776/ub.jkptb.2020.008.02.06.
- [7] Y. Cheng, K. Chen, H. Sun, Y. Zhang, and F. Tao, "Data and knowledge mining with big data towards smart production," *J. Ind. Inf. Integr.*, vol. 9, no. September, pp. 1–13, 2018, doi: 10.1016/j.jii.2017.08.001
- [8] M. Arhami, M. Kom, and S. T. Muhammad Nasir, *Data Mining-Algoritma dan Implementasi*. Penerbit Andi, 2020.
- [9] L. J. Moleong, "Metode Penelitian Kualitatif Cet. 21," *Bandung: Rosda Karya*, 2005.
- [10] C. A. Sugianto, A. H. Rahayu, and A. Gusman,



- “Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah,” *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020, doi: 10.47292/joint.v2i2.30.
- [11] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, “PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda),” *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [12] Kusumadewi dan Purnomo. 2010. Aplikasi Logika Fuzzy untuk mendukung keputusan. Yogyakarta: Graha Ilmu.
- [13] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Education India, 2016.
- [14] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, “kROp: k-Means clustering based routing protocol for opportunistic networks,” *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 4, pp. 1289–1306, 2019.
- [15] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, “K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir,” *IEEE Access*, vol. 8, pp. 151847–151855, 2020.
- [16] S. Muharni, S. Andriyanto, and D. Naista, “Implementasi Dempster Shafer Untuk Mendiagnosa Gangguan Kehamilan Pada Ibu,” *J. Inform.*, vol. 21, no.