

KLASTERISASI DATA REKAM MEDIS PASIEN MENGGUNAKAN METODE K-MEANS CLUSTERING DI RUMAH SAKIT ANWAR MEDIKA BALONG BENDO SIDOARJO

Amir Ali

STIKES Yayasan RS. Dr. Soetomo Surabaya
E-mail : amir.consulting@gmail.com

Abstrak

Penggunaan sistem manajemen informasi yang dimiliki oleh rumah sakit masih terbatas digunakan hanya untuk operasional transaksi pelayanan pasien sehari-hari dan pembuatan laporan saja. Penggunaan SIMRS ini belum optimal, seharusnya tumpukan data yang tersimpan di dalam database server dapat digunakan untuk menghasilkan informasi baru jika kita gali lebih dalam dengan pendekatan IT. Penelitian ini menggunakan teknik data mining dengan metode *K-Means clustering* untuk melakukan klusterisasi data rekam medis pasien. Hasil dari penelitian ini menghasilkan kolom 4 cluster yang terdiri dari kecamatan, diagnosa penyakit, usia dan jenis kelamin. Pada cluster 1 menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin perempuan sebanyak 79(15%) orang, cluster 2 menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin laki-laki sebanyak 214(50%) orang, begitu juga dengan cluster 3 menghasilkan banyak pasien berjenis kelamin perempuan sebanyak 89(17%) orang dan cluster 4 menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin perempuan sebanyak 152(28%) orang. Pengelompokan data rekam medis pasien ini menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit di setiap kecamatan berdasarkan data rekam medis pasien dari rumah sakit anwar medika sebanyak 534 data dengan waktu penyelesaian sebanyak 0.06 detik.

Kata kunci : *K-Means clustering, SIMR, Data Mining, IT*

Abstract

The use of information management systems that are owned by hospitals is still limited to being used only for the operation of daily patient service transactions and making reports only. The use of SIMRS is not optimal, it should pile the data stored in the database server can be used to generate new information if we dig deeper with the IT approach. This study uses data mining techniques with K-Means clustering method to cluster the patient's medical record data. The results of this study produce column 4 clusters consisting of districts, diagnoses of diseases, age and sex. The results of this study produce column 4 clusters consisting of districts, diagnoses of diseases, age and sex. Cluster 1 produced many patients consisting of 79(15%) female patients, Cluster 2 produced many patients consisting of 214(50%) male patients. Likewise Cluster 3 produced 89(17%) female patients. people and cluster 4 produced many patients consisting of 152(28%) female patients. The grouping of patient medical record data produces new information about the pattern of grouping of disease spread in each district based on the patient's medical record data from Anwar Medika Hospital as much as 534 data with a completion time of 0.06 seconds

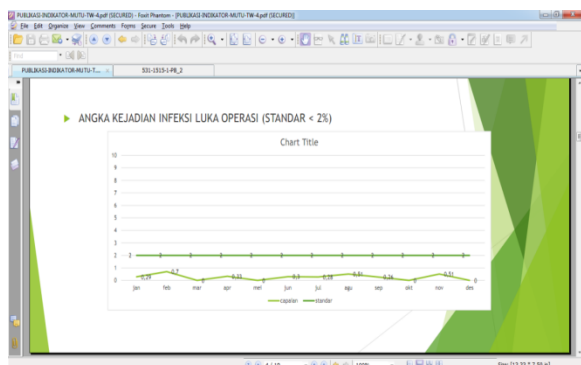
Keyword : *K-Means clustering, SIMR, Data Mining, IT*

I. PENDAHULUAN

Data mining merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai database yang besar [1]. Salah satu fungsi data mining adalah clustering.

Clustering merupakan teknik pengelompokan record pada basis data berdasarkan kriteria tertentu. Hasil clustering diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data [2]. Konsep dasar dari clustering adalah mengelompokkan sejumlah objek ke dalam cluster dimana cluster yang baik adalah cluster yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu cluster dan tingkat ketidaksamaan yang tinggi dengan objek cluster yang lainnya. Metode clustering hadir dikarenakan kondisi berlimpahnya data yang merupakan akumulasi data transaksi yang terekam selama bertahun-tahun [3].

Data pasien yang tersimpan pada database aplikasi SIMRS di rumah sakit pada umumnya hanya dimanfaatkan untuk membuat laporan dan grafik pasien rumah sakit, data penyakit pasien serta biaya berobat pasien. Database yang ada belum dapat digunakan sebagai dasar dalam pengambilan keputusan.



Gambar 1. Angka kejadian Infeksi Luka Operasi

Gambar diatas merupakan laporan pada area klinis yang dibuat oleh petugas kesehatan rumah sakit anwar medika sidoarjo. Pada laporan ini hanya disampaikan informasi angka capaian dan nilai standar yang dicapai tetapi jika digali lebih dalam maka dapat diinformasikan bahwa terdapat

bulan-bulan yang nilai/ angka capaiannya bernilai 0 yang mengelompok yang mana hal ini tidak boleh terjadi karena tidak mencapai nilai standar yang ditentukan. Hal ini dapat kita analisa lebih dalam menggunakan teknik data mining dengan metode K-Means clustering, agar didapatkan informasi baru dari data klinis pasien yang tersimpan dalam database rumah sakit.

Oleh karena itu diperlukan pendekatan IT yang menggunakan teknik data mining dengan metode k-means karena dalam proses analisis clustering metode yang digunakan untuk membagi data menjadi subset data berdasarkan kesamaan atau kemiripan yang telah ditentukan sebelumnya. Sehingga dalam menemukan informasi baru berdasarkan data rekam medis pasien yang tersimpan dalam database SIMRS (Sistem Informasi Manajemen Rumah Sakit) Anwar Medika Sidoarjo agar dapat digunakan dalam proses pengambilan keputusan oleh pihak manajemen rumah sakit agar tepat sasaran.

II. METODOLOGI

Clustering

Clustering atau pengklasteran adalah suatu teknik data mining yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokan data atau lebih tepatnya mempartisi dari dataset ke dalam subset. Teknik cluster mempunyai dua metode dalam pengelompokannya yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Non-hierarchical clustering* pada teknik ini dimulai dengan menentukan jumlah cluster yang diinginkan (dua cluster, tiga cluster, empat cluster atau lebih), setelah jumlah yang cluster yang diinginkan maka proses cluster dimulai tanpa mengikuti proses hirarki, metode ini juga sering disebut sebagai metode K-Means Clustering [4]

K-Means Clustering

Metode K-Means Clustering merupakan metode yang paling populer dan banyak digunakan pada proses klusterisasi data. Metode K-Means Clustering membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas [5]

Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam

satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

KMeans Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [6].

K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara partisi yang memisahkan data ke dalam kelompok yang berbeda – berda. Algoritma ini mampu meminimalkan jarak antara data ke clusternya. Pada dasarnya penggunaan algoritma ini dalam proses clustering tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses. Sehingga dalam penggunaan algoritma kmeans terdapat aturan sebagai berikut [7]: a. Berapa jumlah cluster yang perlu dimasukkan dan b. Hanya memiliki atribut bertipe numerik

Pada dasarnya algoritma k-means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat cluster awal, pada penentuan pusat cluster ini dipilih secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing – masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat cluster yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat cluster. Selanjutnya posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap – tiap cluster dan terakhir akan terbentuk cluster baru.

Algoritma K-Means pada dasarnya melakukan 2 proses yakni proses pendeteksian okasi pusat cluster dan proses pencarian anggota dari tiap-tiap cluster. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster.

Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean. Suatu data akan menjadi anggota dari cluster ke- k apabila jarak data tersebut ke pusat cluster ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lain

Euclidian Distance

Dalam mengukur jarak antara titik pusat atau *centroid* dengan data adalah dengan menggunakan *euclidean distance*. Rumus *euclidean distance* adalah sebagai berikut (Fayyad, 1996) :

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

Dengan:

d_{ik} = jarak antara data ke- i dengan titik pusat kluster ke- k

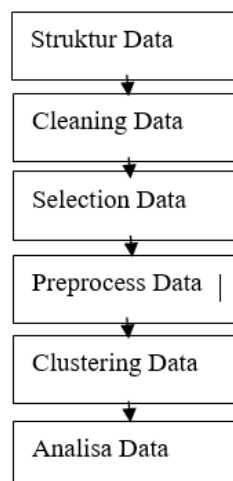
m = jumlah atribut

x_j = data ke- i

c_k = data pusat kluster ke- k

III. HASIL DAN PEMBAHASAN

Data pada penelitian ini berjumlah 534 data selama triwulan pertama di tahun 2019 pada bulan januari, februari, maret. Atribut yang digunakan dalam proses clustering adalah atribut jenis kelamin, usia, kecamatan dan diagnosa penyakit. Berikut tahapan dalam penelitian ini :



Gambar 2. Tahapan Penelitian[8]

Tahapan penelitian ini dimulai dari tahapan sebagai berikut :

1. Struktur Data

Pengambilan data rekam medis dari RS Anwar Medika berupa file excel yang berisi yaitu noRM,

nama pasien, jenis kelamin, umur, alamat, diagnosa penyakit

2. Cleaning Data

Dari 600 data akan dilakukan pembersihan data, dimana informasi yang kurang jelas dan tidak lengkap akan dibuang. Seperti informasi tidak jelasnya informasi alamat dan data yang berada diluar kecamatan Sidoarjo.. Dari 600 data didapatkan data bersih sebanyak 534

3. Selection Data

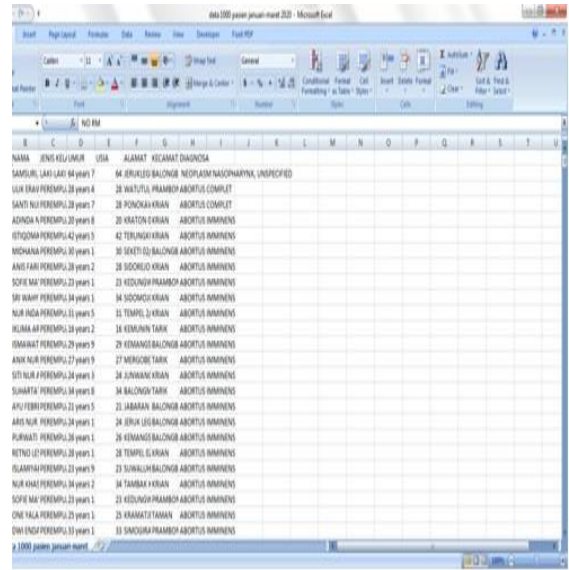
Informasi data-data yang sudah lengkap dalam setiap atribut akan dilakukan seleksi. Atribut yang dilakukan seleksi yaitu, noRM, nama, jenis kelamin, usia, alamat, diagnosa. Alamat ini akan ditransformasikan ke dalam kecamatan yang memuat alamat rumah pasien tersebut.

NAMA	JENIS KELAMIN	USIA	ALAMAT	KECAMATAN	DIAGNOSA
SAMUEL LAJI LAJI	84 years 7	64	ARUKES BALONG	NEOPULAN	ABORTUS INMANSUS
USUR RABU PEREMPU	38 years 4	28	WATULU PRAMBON	ABORTUS COMPLET	
SANTI NU PEREMPU	38 years 8	28	PONOKAH KRAN	ABORTUS COMPLET	
ADINDA A PEREMPU	38 years 8	28	KRATON KRAN	ABORTUS INMANSUS	
ISTIQOMAH PEREMPU	42 years 8	42	TERUNGAN KRAN	ABORTUS INMANSUS	
MICHANA PEREMPU	38 years 1	38	SEKATI (2) BALONG	ABORTUS INMANSUS	
ANIS HARI PEREMPU	38 years 2	38	SOEKILO KRAN	ABORTUS INMANSUS	
SCHIE HAR PEREMPU	38 years 2	38	KEDUNGGU PRAMBON	ABORTUS INMANSUS	
SRI HARIP PEREMPU	38 years 1	38	SOEKILO KRAN	ABORTUS INMANSUS	
NAUR NELLA PEREMPU	38 years 1	38	TEMPEL 2 KRAN	ABORTUS INMANSUS	
WILMA A PEREMPU	38 years 1	38	KEDUNGGU TARK	ABORTUS INMANSUS	
EMANAT PEREMPU	38 years 1	38	KEMANGS BALONG	ABORTUS INMANSUS	
ANUR NUR PEREMPU	27 years 9	27	MENGENG TARK	ABORTUS INMANSUS	
SITI NUR PEREMPU	38 years 1	38	JURANGAN KRAN	ABORTUS INMANSUS	
SUMARTI PEREMPU	38 years 8	38	BALONG TARK	ABORTUS INMANSUS	
AFU FERDI PEREMPU	38 years 1	38	JABARAN BALONG	ABORTUS INMANSUS	
ANUR NUR PEREMPU	38 years 1	38	ERUK (2) BALONG	ABORTUS INMANSUS	
PURNATI PEREMPU	38 years 1	38	KEMANGS BALONG	ABORTUS INMANSUS	
WETNO (2) PEREMPU	38 years 1	38	TEMPEL KRAN	ABORTUS INMANSUS	
ELAMAH PEREMPU	38 years 1	38	SUNWAL BALONG	ABORTUS INMANSUS	
NUR KHAI PEREMPU	38 years 2	38	TAMBA KRAN	ABORTUS INMANSUS	
SOPE MA PEREMPU	38 years 1	38	KEDUNGGU PRAMBON	ABORTUS INMANSUS	
ONE YALA PEREMPU	38 years 1	38	KRAMAT TAMAN	ABORTUS INMANSUS	
OWI ENDU PEREMPU	38 years 1	38	SANDUHA PRAMBON	ABORTUS INMANSUS	

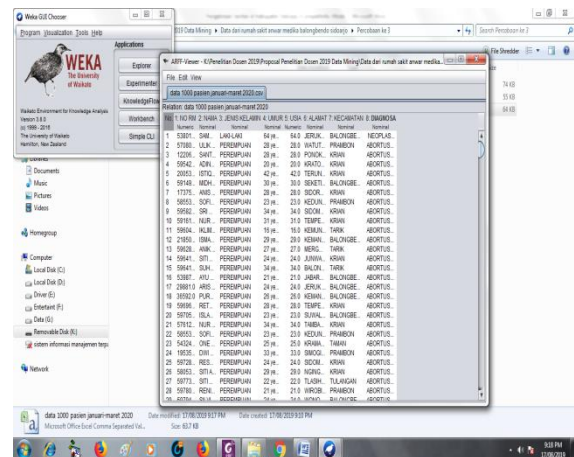
Gambar 3. Seleksi data yang kurang lengkap informasinya

4. Preprocess Data

Proses selanjutnya adalah menyimpan data excel dalam format bentuk data .csv. Preprocess dilakukan pada aplikasi weka dengan memasukkan data yang diseleksi, kemudian data-data tersebut disimpan dalam bentuk data .arff agar dapat dibaca oleh aplikasi weka. Pada tahap ini atribut yang dipilih yaitu atribut umur, diagnosa penyakit dan kecamatan.

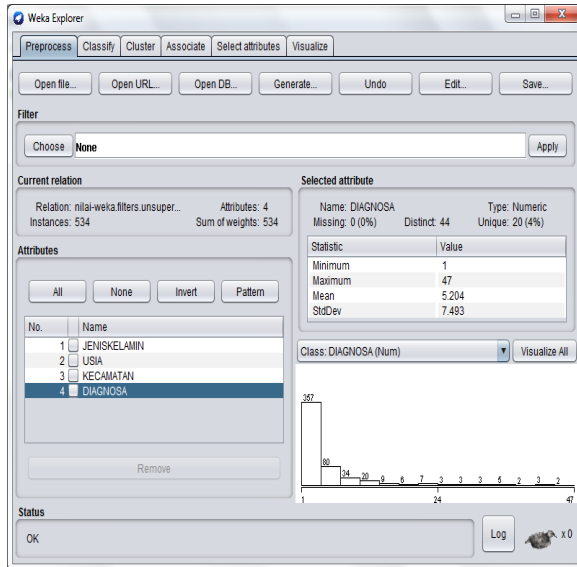


Gambar 4. Data dalam bentuk format ekstensi .CSV



Gambar 5. Mengubah data ke dalam format file ekstensi .Arff

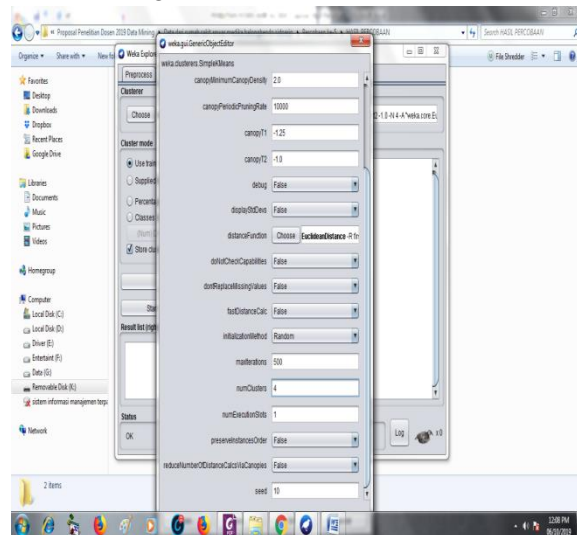
Pada gambar 5 data diubah ke dalam format file ekstensi .Arff agar system dapat membaca datanya dari format file .csv di microsoft excel.



Gambar 6. Atribut yang digunakan dalam proses clustering

Pada gambar 6 diatas, data dalam format .Arff dapat dapat dibaca sehingga data tampak visualisasinya seperti diatas, dengan atribut jenis kelamin, usia, kecamatan dan diagnosa penyakitnya

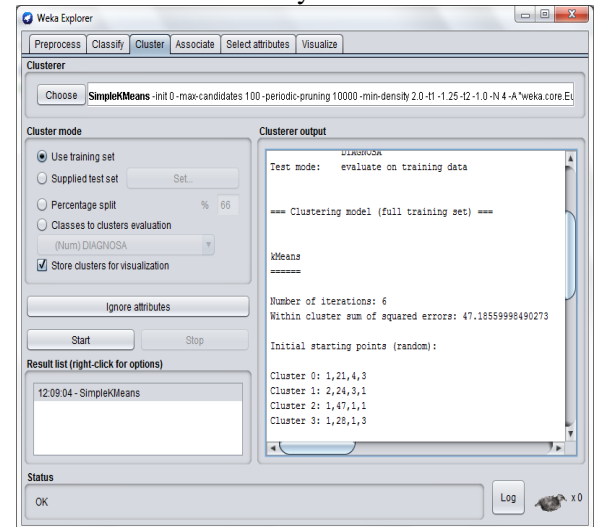
5. Clustering Data



Gambar 7. Menentukan Jumlah Cluster Yang Akan dibentuk

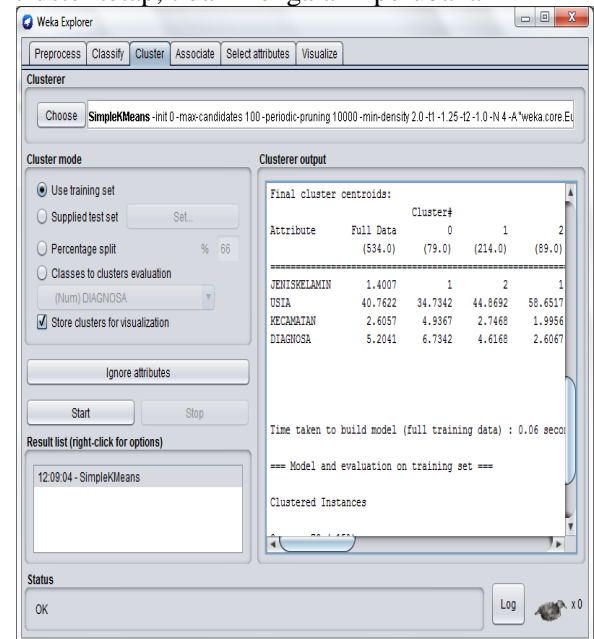
Pada tahapan seperti gambar 7 diatas ini akan dilakukan pengelompokan data berdasarkan kesamaan ciri/ karakteristik yang sama dalam satu cluster. Pengelompokan data ini

menggunakan algoritma K-Means. Jumlah cluster ditentukan sebanyak 4 cluster.

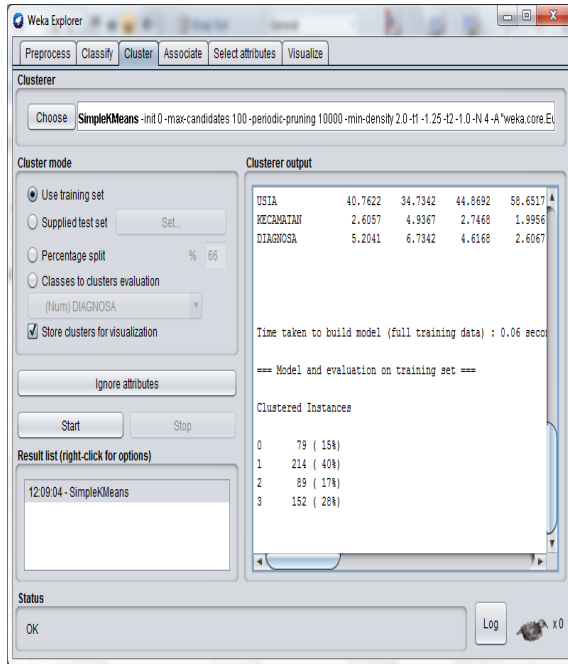


Gambar 8. Proses Clustering berhenti pada iterasi ke-6

Pada proses clustering, iterasi berhenti pada iterasi ke-6, dimana anggota dari masing-masing cluster tetap, tidak mengalami perubahan

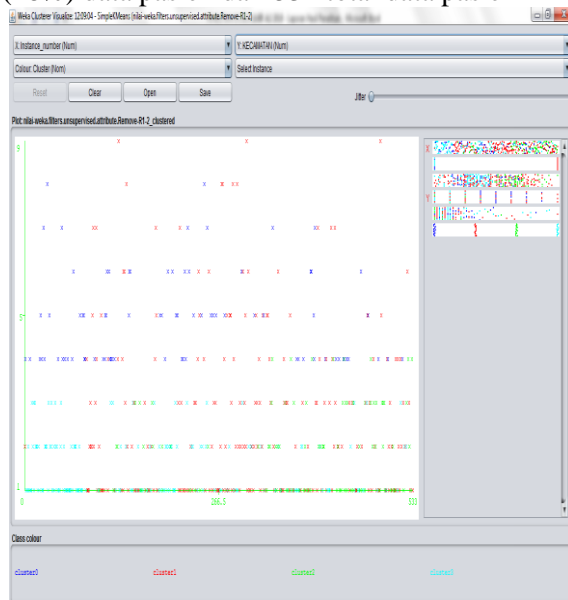


Gambar 9. Proses Final Clustering



Gambar 10. Proses Final Clustering

Tampak dari gambar 10 pada proses final clustering, didapatkan hasil bahwa terdapat data sebanyak 79 (15%) data pasien dari 534 total data pasien yang menempati cluster 0. data sebanyak 214 (40%) data pasien dari 534 total data pasien, data sebanyak 89 (17%) data pasien dari 534 total data pasien dan data sebanyak 152 (28%) data pasien dari 534 total data pasien



Gambar 11. Visualisasi Clustering Data Pasien

6. Analisa Data

Dengan menggunakan algoritma K-Means Clustering dimana akan ditentukan 4 cluster untuk 4 atribut yaitu jenis kelamin, usia, kecamatan dan diagnosa penyakit

Dari informasi pada tabel diatas, maka didapatkan hasil akhir seperti tabel berikut :

Tabel 1. Hasil Clustering pada cluster 1 dan cluster 2

Hasil Cluster 1	Hasil Cluster 2
<p>Terdiri dari 79 Orang yang berasal dari,</p> <p>Kecamatan :</p> <ol style="list-style-type: none"> 1. Prambon = 36 2. Taman = 22 3. Wonoayu = 13 4. Tulangan = 6 5. Sukodono = 2 	<p>Terdiri dari 214 Orang yang berasal dari,</p> <p>Kecamatan :</p> <ol style="list-style-type: none"> 1. Krian = 81 2. Balongbendo = 37 3. Tarik = 34 4. Prambon = 21 5. Taman = 16 6. Wonoayu = 8 7. Tulangan = 7 8. Sukodono = 5 9. Waru = 3 10. Tanggulangin = 2
<p>Diagnosa Penyakit</p> <ol style="list-style-type: none"> 1. Anomali refraksi = 20 2. Animal bite,vulnus apertum tdk ditentukan = 7 3. Abortus imminens = 9 4. Abses = 6 5. Abortus incompletet us = 7 6. Amenore sekunder = 5 7. ANEMIA = 5 8. ANAL ABCESS = 2 9. AECB(CHR ONIC OBSTRUC 	<p>Diagnosa Penyakit :</p> <ol style="list-style-type: none"> 1. Anomali refraksi = 81 2. Animal bite,vulnus apertum tdk ditentukan = 60 3. Allergic contact dermatitis, unspecified cause = 22 4. Anemia = 8 5. Anal abscess = 6 6. Aecb(chronic obstructive pulmonary disease with acute exacerbation) = 4 7. Acute appendicitis, unspecified = 3 8. Angina stabil = 2 9. Acute transmural myocardial infarction of unspecified site = 4 10. Anemia gravis = 2 11. Anemia aplastik = 1

Hasil Cluster 1	Hasil Cluster 2
TIVE PULMONARY DISEASE WITH ACUTE EXACERBATION) = 3	12. Ankylosing spondylitis = 1
10. ANGINA STABIL = 1	13. Alo;lbbb;aki;hyper kalemia = 1
11. AMENORR HEA = 2	14. Alo;dc (decompensasi cordis);pneumonia = 1
12. ABSCESS OF BARTHOLIN'S GLAND = 4	15. Alo;dc (decompensasi cordis);ckd (chronic kidney disease) = 1
13. ANGIOEDEMA = 1	16. Acute transmural myocardial infarction of unspecified site ;ima (infrak myokard akut) = 1
14. ABORTUS COMPLET = 1	17. Acute bronchiolitis, unspecified ;demam dengue = 1
15. ANEMIA KEHAMILAN (NIFAS) = 1	18. Abses cerebri = 1
16. ANEMIA GRAVIS = 1	19. Neoplasm nasopharynx, unspecified = 1
17. ACQUIRED ABSENCE OF PART OF HEAD AND NECK = 1	
18. ACNE VULGARIS = 2	
19. ABSES PARU;TUBERKULOSIS PARU LAINNYA = 1	
Usia : 1. 1-25 = 27 2. 26 - 49 = 36 3. ≥ 50 = 1	Usia : 1. 1-25 = 43 2. 26 - 49 = 83 3. ≥ 50 = 88

Hasil Cluster 1	Hasil Cluster 2
Jenis Kelamin : 1. Laki-Laki = 0 2. Perempuan = 79	Jenis Kelamin : 1. Laki-Laki = 214 2. Perempuan = 0

Tabel 2. Hasil Clustering pada cluster 3 dan cluster 4

Hasil Cluster 3	Hasil Cluster 4
Terdiri dari 89 Orang yang berasal dari, Kecamatan : 1. Krian = 36 2. Balongbendo = 25 3. Tarik = 19 4. Prambon = 8 5. Tulangan = 1	Terdiri dari 152 Orang yang berasal dari, Kecamatan : 1. Krian = 86 2. Balongbendo = 48 3. Tarik = 18
Diagnosa Penyakit : 1. Anomali refraksi = 67 2. Animal bite,vulnus apertum tdk ditentukan = 3 3. Allergic contact dermatitis, unspecified cause = 8 4. Abses = 1 5. Amenore sekunder = 3 6. Anemia = 2 7. Acute appendicitis , unspecified = 2 8. Angioneurotic odema = 1 9. Acquired absence of part of head and neck = 1	Diagnosa Penyakit : 1. Anomali refraksi = 31 2. Animal bite,vulnus apertum tdk ditentukan = 14 3. Abortus imminens = 38 4. ALLERGIC CONTACT DERMATITIS, UNSPECIFIED CAUSE = 7 5. Abses = 2 6. Abortus incompletetus = 17 7. Amenore sekunder = 17 8. ANEMIA = 3 9. AECB(CHRONIC OBSTRUCTIVE PULMONARY DISEASE WITH ACUTE EXACERBATION) = 1 10. ACUTE APPENDICITIS,

Hasil Cluster 3	Hasil Cluster 4
10. Anemia gravis;hypokalemia = 1	UNSPECIFIED = 2
	11. ANGINA STABIL =2
	12. AMENORRHEA =2
	13. ANGINEUROTTIC OEDEMA = 1
	14. ABORTUS COMPLET = 1
	15. ANEMIA KEHAMILAN (NIFAS) = 1
	16. ANEMIA APLASTIK = 1
	17. ABSSES PAYUDARA = 1
	18. ANGINA PECTORIS, UNSPECIFIED =1
	19. AMENORE SEKUNDER;AMENORE SEKUNDER = 1
	20. AIHA (AUTO IMMUNE HEMATOMA ANEMI);ANEMIA GRAVIS;HYPOGLIKECEMIA = 1
	21. AECB(CHRONIC OBSTRUCTIVE PULMONARY DISEASE WITH ACUTE EXACERBATION);HHF (HYPERTENSI FAILURE) = 1
	22. ADHESIVE CAPSULITIS OF SHOULDER = 1
	23. ACUTE BRONCHIOLITIS,

Hasil Cluster 3	Hasil Cluster 4
	UNSPECIFIED ;ACUTE BRONCHIOLITIS, UNSPECIFIED =2
	24. ABSCESS, FURUNCLE AND CARBUNCLE OF NOSE =1
	25. ABSCESS BUCCAL CAVITY = 1
	26. ABORTUS PROVOKATUS =1
	27. ABORTUS INCOMPLETE TUS;CURETAGGE = 1
Usia :	Usia :
1. 1-25 = 0	1. 1-25 = 60
2. 26 – 49 = 14	2. 26 – 49 = 91
3. ≥ 50 = 75	3. ≥ 50 = 1
Jenis Kelamin :	Jenis Kelamin :
1. Laki-Laki = 0	1. Laki-Laki = 0
2. Perempuan = 89	2. Perempuan = 152

Dari data tabel 1 dan tabel 2 diatas dapat menghasilkan kolom 4 cluster yang terdiri dari kecamatan, diagnosa penyakit, usia dan jenis kelamin. Pada cluster yang pertama menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin perempuan sebanyak 79 orang, cluster dua menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin laki-laki sebanyak 214. begitu juga dengan cluster tiga menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin perempuan sebanyak 89 orang, dan cluster empat menghasilkan banyak pasien yang terdiri dari pasien yang berjenis kelamin perempuan sebanyak 152 orang.

Pada penelitian ini peneliti, menggunakan 6 atribut yang dibutuhkan dalam proses clustering yaitu NoRM, NamaPasien, Jenis kelamin, Usia ,

Alamat dan Diagnosa. Seperti pada tabel berikut ini :

Tabel 3. Data Pasien Sesuai Atribut RS Anwar Medika Balongbendo Sidoarjo

No RM	Nama Pasien	Jenis Kelamin	Usia	Alamat	Diagnosa
5380 XX	SAMSURI, TN	LAKI - LAKI	64	JERUKLEGI 15/04 BALONGBENDO	NEOP LASM NASO PHARYNX, UNSPECIFIED
No RM	Nama Pasien	Jenis Kelamin	Umur	Alamat	Diagnosa
5708 XX	ULIKERAWATI, NY	PEREMPURAN	28	WATUTULIS 01/06 PRAMBON	ABORTUS COMPLET
1220 XX	SANTINUR AULIYA, NY	PEREMPURAN	28	PONOKAWAN 07/03 KRIAN	ABORTUS COMPLET
5954 XX	ADINDA NOVI ANTIKA I, NY	PEREMPURAN	20	KRATON 011/03 KRIAN	ABORTUS IMMINENS
.....
.....
...				

Untuk transformasi data dapat dilakukan dengan langkah-langkah sebagai berikut :

- Mengurutkan data berdasarkan frekuensi kemunculannya
- Inisialisasikan data tersebut mulai dari data tertinggi dengan nilai 1, kemudian data selanjutnya 2,3 dan seterusnya

Contoh transformasi variabel kecamatan pada tabel di bawah ini :

Tabel 4. Transformasi variabel kecamatan

Kecamatan	Jumlah	Inisial
KRIAN	200	1
BALONGBENDO	110	2
TARIK	73	3
PRAMBON	66	4
TAMAN	39	5
WONOAYU	21	6
TULANGAN	11	7
SUKODONO	7	8
WARU	3	9
TANGGULANGIN	2	10

Setelah dilakukan transformasi data selanjutnya adalah melakukan perhitungan setiap data ke setiap titik pusat cluster untuk iterasi 1 adalah sebagai berikut :

- Menentukan terlebih dahulu pusat awal cluster

Tabel 5. Titik Pusat Awal Cluster

Pusat Awal Cluster	Jenis Kelamin	Usia	Kecamatan	Diagnosa Penyakit
C0	2	74	1	5
C1	1	42	3	16
C2	2	18	5	2
C3	1	16	2	1

- Menghitung jarak dari pusat cluster ke masing-masing data

- Hitung jarak data ke-1 ke pusat cluster pertama C0 =

$$\sqrt{(2 - 2)^2 + (74 - 64)^2 + (1 - 2)^2 + (5 - 47)^2} = 43,19$$

- Hitung jarak data ke-1 ke pusat cluster kedua C1 =

$$\sqrt{(1 - 2)^2 + (42 - 64)^2 + (3 - 2)^2 + (16 - 47)^2} = 38,04$$

- Hitung jarak data ke-1 ke pusat cluster ketiga C2 =

$$\sqrt{(2 - 2)^2 + (18 - 64)^2 + (5 - 2)^2 + (2 - 47)^2} = 64,42$$

- d. Hitung jarak data ke-1 ke pusat cluster keempat C3 =

$$\sqrt{(1-2)^2 + (16-64)^2 + (2-2)^2 + (1-47)^2} = 66,50$$

Hasil Clustering

Dari perhitungan diatas, didapatkan hasil sebagai berikut :

Tabel 6. Hasil Perhitungan Iterasi Pertama

No	Jenis Kelamin	Usia	Kecamatan	Diagnosa Penyakit	Jarak Ke			
					C0	C1	C2	C3
1	2	64	2	47	43.19	38.04	64.42	66.49

Dari tabel diatas, untuk data ke 1 setelah dilakukan perhitungan jarak maka data ke 1 ini jaraknya lebih mendekati cluster 1, sehingga data ke 1 ini menjadi anggota dari cluster 1. Hal ini sesuai dengan metode *K-Means Clustering* dalam menganalisis data rekam medis pasien. Adapun tahap analisis data sebagai berikut :

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

Dari pentahapan diatas pada point 4, data akan dikelompokkan ke dalam cluster berdasarkan jarak terdekat.

IV. SIMPULAN DAN SARAN

Kesimpulan

Dari hasil penelitian ini, dapat disimpulkan bahwa :

- 1) Peneliti dapat mengidentifikasi data rekam medis dari rumah sakit anwar medika sebanyak 534 data pasien dengan waktu penyelesaian sebanyak 0.06 detik oleh sistem
- 2) Peneliti dapat menganalisa data rekam medis untuk dikelompokkan menjadi 4

cluster dengan 4 variabel yaitu variabel kecamatan, diagnosa penyakit, usia dan jenis kelamin.

- 3) Pengelompokan data rekam medis pasien dari proses data mining diatas adalah untuk menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit di setiap kecamatan. Hal ini dapat dijadikan acuan bagi pemerintah daerah terutama dinas kesehatan kabupaten sidoarjo untuk melakukan sosialisasi dan pencegahan terhadap sumber penyakit berdasarkan diagnosa penyakit, usia dan jenis kelamin yang banyak diderita pasien.

Saran

Agar ke depan penelitian ini dapat dikembangkan dengan variasi data yang lebih beragam dan metode clustering yang berbeda

V. UCAPAN TERIMA KASIH

Terima kasih kami sampaikan kepada kampus Stikes Yayasan Rumah Sakit Dr. Soetomo yang telah mendanai penelitian ini, sehingga dapat terlaksana dengan baik dan selesai. Terima kasih juga kami ucapkan kepada teman-teman dosen yang telah membantu dalam menyelesaikan penelitian ini. Semoga penelitian ini dapat membawa bermanfaat yang besar

REFERENSI

- [1] L. Daniel, *Discovery Knowledge in Data. Canada*. Kanada: A Jhon Wiley & Sons, Inc Publication, 2015.
- [2] Jiawei Han and Micheline Kember, *Data Mining: Concepts and Techniques Second Edition*. San Francisco, USA: Morgan Kaufmann Publishers, 2006.
- [3] I. S. M. Silitonga, D Parasian, "Klusterisasi pola penyebaran penyakit pasien berdasarkan usia pasien dengan menggunakan K-Means clustering.," *Times Technol. informatics Comput. Syst.*, vol. VI No.2. 2, 2017.
- [4] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *J. Ilm. Tek. Ind.*, vol. 12 No 1, pp. 10-20, 2013.
- [5] B. Pavel, *Survey of Clustering Data*

- Mining Techniques*. 2002.
- [6] Y. Agusta, "K-means - Penerapan, Permasalahan dan Metode Terkait," *Sist. dan Inform.*, vol. 3, pp. 47–60, 2007.
- [7] A. T. dan F. K. S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," 2012.
- [8] G. Wardani, Ni Wayan. Murni, Ni Nyoman. Luka, Siti Saibah Pua. Indrawan, "Analisis Penerapan K-means Untuk Pengelompokan Diagnosa Penyakit Kulit dan Kelamin Berdasarkan Rentang Usia," *Pros. Semin. Nas. Pendidik. Tek. Inform. (Senapati 2016)*, 2016.