

Implementasi *Data Mining* Algoritma *K-Means* untuk *Clustering* Penyakit di RS Panti Waluyo Surakarta

¹Anggit Hidayah*, ²Wiji Lestari, ³Eko Purwanto

¹Program S1-Sistem Informasi, Universitas Duta Bangsa, Surakarta, Indonesia

^{2,3}Universitas Duta Bangsa, Surakarta, Indonesia

*anggit.nurhidayah@fikom.udb.ac.id

ABSTRAK

Kebutuhan informasi dan pengetahuan baru untuk pendukung keputusan Rumah Sakit sangat diperlukan, ancaman gangguan kesehatan mulai diutarakan oleh badan kesehatan dunia dan menjadi fokus pemerintahan dalam usaha peningkatan kesejahteraan kesehatan masyarakat. Untuk itu diperlukan pengelompokan penyakit untuk mengetahui pola / jenis gangguan kesehatan dengan jumlah banyak maupun sedikit. Dikarenakan jumlah data rekam medis dan variabel data yang banyak maka dibutuhkan metode untuk mempermudah pengelompokan penyakit. Dengan pendekatan pengklasteran K-means, pembagian kelompok penyakit dapat dilakukan berdasarkan 3 variabel yaitu umur (vu), kode penyakit (vi) dan kecamatan (vk). Pada penelitian ini dilakukan pengklasteran menggunakan algoritma K-means yang diimplementasikan kedalam kode bahasa pemrograman. Hasil dari pengelompokan direpresentasikan dalam bentuk grafik sehingga pihak eksekutif lebih mudah memahami hasil pengelompokan data tersebut. Dari proses pengelompokan 200 data rekam medis periode rebruari 2019 didapatkan 3 kelompok yaitu cluster 1 dengan penyakit tertinggi Chronic ischaemic heart disease.

Kata Kunci: Penyakit, Data mining, Algoritma K-means

Latar Belakang

Kesehatan merupakan salah satu aspek penting yang didorong oleh pemerintah untuk menjadi salah satu isu nasional selain peningkatan pembangunan infrastruktur, data rekam medis digunakan sebagai salah satu acuan pengukuran tingkat penyebaran penyakit yang diderita oleh masyarakat di suatu wilayah. Kesehatan keluarga adalah suatu kondisi yang sempurna baik secara fisik, mental maupun sosial serta mampu untuk memelihara kehidupannya sendiri dan kehidupan (Eliana dan Sri Sumiati, 2016).

Organisasi Kesehatan Dunia atau World Health Organization (WHO) pada tahun 2019 menyampaikan informasi ancaman kesehatan global terbesar yang dapat mengancam jutaan nyawa diseluruh dunia, 10 ancaman kesehatan tersebut yaitu penolakan vaksin, resistensi obat, polusi udara dan perubahan iklim, pandemi flu global, krisis pada wilayah rentan, ancaman patogen, penyakit tidak

menular, Demam Berdarah Dengue, HIV dan layanan kesehatan yang lemah.

Perubahan iklim tidak hanya berdampak pada lingkungan, namun juga berdampak darurat medis terutama masyarakat yang mempunyai masalah kardiovaskular, diabetes, dan pernafasan akan lebih rentan terhadap penyakit (Yustiningsih, 2019). Tingkat layanan kesehatan juga berpengaruh pada kesiapan dan kesuksesan menangani ancaman global yang sewaktu-waktu dapat terjadi. Paradigma pengelolaan manajemen data pencatatan riwayat pasien berobat sudah mulai dilakukan pada rumah sakit yang sudah mempunyai sistem informasi dengan pengelolaan mandiri. Digitalisasi era 4.0 sudah merambah ke dunia kesehatan bukan hanya untuk operasional namun saat ini telah masuk ke pengolahan data rekam medis dan analisis kesehatan (Adikrishna dan Mandira, 2019).

Metodologi Penelitian

Metode pengembangan sistem menggunakan metode prototyping. *Prototyping* adalah metode untuk mengembangkan sebuah sistem yang didalamnya banyak dilakukan interaksi antara pengembang dengan pengguna selama proses pembuatan sistem (Rudiansyah, 2018).

1. *Communication* / Komunikasi Tim Pengembang

Pada tahap ini penulis melakukan wawancara dengan pihak rumah sakit mengenai kebutuhan data yang akan diolah beserta atribut database yang akan dijadikan variable perhitungan data mining.

2. *Quick Plan* / Perencanaan Secara Cepat Sederhana

Penulis melakukan desain interface sistem yang akan dibuat dan menu apa saja yang akan dijadikan tampilan pada sistem. Pembuatan prototype program sederhana yang dapat menggambarkan secara garis besar.

3. *Modeling Quick Design* / Model Rancangan

Dilakukan pemodelan UML yaitu *Use Case diagram* untuk mendefinisikan fungsi sistem dan *Activity Diagram* untuk menggambarkan alur proses bisnis, rancangan navigasi menu dll.

4. *Construction of Prototype* / Pembuatan Prototype

Impelementasi dari rancangan kedalam bentuk sistem informasi berbasis website dengan menu-menu sesuai yang telah ditetapkan pada tahap sebelumnya.

5. *Deployment Delevery & Feedback* / Penyerahan

Setelah sistem sudah menjadi sebuah suatu aplikasi yang siap pakai, maka akan dilakukan pengetesan terlebih dahulu sebelum digunakan. Pengujian dilakukan dari berbagai aspek diantaranya hasil cluster yang dihasilkan, integritas perhitungan data. Memberikan umpan balik dan evaluasi terhadap sistem yang sudah di buat dan Tim pengemabng melakukan perbaikan terhadap prototype.

Hasil dan Pembahasan

Agar data dapat diolah menggunakan algoritma K-means *clustering*, maka jenis data kode *ICD-10* dan kecamatan harus di transformasikan ke dalam bentuk angka.

Tabel 1. Inisiasi penyakit

Kode / penyakit	Inisiasi
Amoebiasis	18
Anemia	24
Demam tifoid dan paratifoid	16
Diare dan gastroenteritis	21
Campak	8
TB paru TBA	19
TB selain paru	17
Kusta	20
Tetanus obstetric	15
Scabies	22

Tabel 2. Inisiasi kecamatan

Kecamatan	Inisiasi
Banjarsari	1
Jebres	2
Laweyan	3
Pasar kliwon	4
Serengan	5
Kartasura	6
Gilingan	8
Nusukan	9
Cemani	11
Colomadu	21

Tabel 3. Inisiasi umur

Rentang umur	Inisiasi
Balita (0-5 tahun)	1
Kanak-kanak (6-11 tahun)	2
Remaja (12-25 tahun)	3
Dewasa (26-45 tahun)	4
Lansia (46-65 tahun)	5
Manula (65-100 tahun)	6

Tabel 4. Data Pasien Rumah Sakit

Data	Umur	Kode Diagnosa	Kecamatan
1	3	16	3
2	3	16	5
3	3	18	3
4	2	18	2
5	3	21	3
6	3	21	3
7	2	21	5
8	3	21	1
9	2	21	4
10	3	21	5
11	1	21	1
12	3	19	4
13	2	17	8
14	5	20	8
15	3	20	9
16	3	20	6
17	4	15	6
18	3	22	11
19	3	24	21
20	2	23	8

Setelah semua data di tranformasikan semuain kebutuhan dan berupa data angka, maka kemudian dapat dilakukan pengelompokan

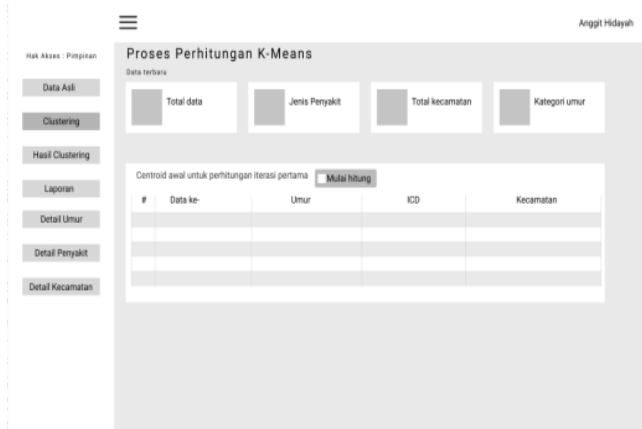
menggunakan algoritma *K- Means Clustering*. Langkah yang perlu dilakukan agar menjadikan beberapa *cluster* yaitu :

- a. Menentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data- data yang akan dikelompokkan menjadi 3 *cluster*.
- b. Menentukan pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal cluster ditentukan secara random dan didapat titik pusat dari setiap *cluster* seperti pada tabel 5 dan contoh data sampel yang ada pada tabel 1.

Tabel 5 Titik pusat *cluster*

Titik pusat awal	Umur	Kode diagnosa	Kecamatan
<i>Centroid 1</i>	2	18	2
<i>Centroid 2</i>	1	21	1
<i>Centroid 3</i>	5	20	8

Halaman dashboard berisi informasi jumlah data pasien yang akan dilakukan *data mining*, jumlah jenis penyakit dan tanggal terakhir kali data tersebut di update / dilakukan perubahan data serta tabel data pasien sebelum dilakukan proses *data mining*. Pada bagian kiri terdapat navbar yang menampilkan menu-menu yang dapat diakses, bagian atas kanan terdapat informasi pengguna yang saat ini digunakan.



Gambar 1. Halaman *Clustering Data*

Halaman proses clustering menunjukkan tabel centroid awal yang digunakan sebagai perhitungan iterasi pertama, kemudian tombol mulai hitung.

Proses Perhitungan K-Means

Data Terbaru

- Total Data: 200
- Jenis Penyakit: 67
- Total Kecamatan: 6
- Kategori Umur: 6

Centroid awal untuk perhitungan iterasi pertama Mulai Hitung

Data ke -	Centroid	Umur (Tahun)	ICD (Kode ICD 10)	Kecamatan (ID Kecamatan)
18	1	6	47	3
35	2	6	56	4
22	3	5	50	1

Gambar 2. Halaman Perhitungan Clustering

Hasil Perhitungan K-Means

Centroid perhitungan terakhir

Centroid ke-	Umur (Tahun)	ICD (Kode Penyakit ICD 10)	Kecamatan (ID Kecamatan)
1	4.77333	44.14667	2.40000
2	4.02041	117.00000	2.71429
3	4.68421	78.76316	2.31579

Hasil perhitungan terakhir

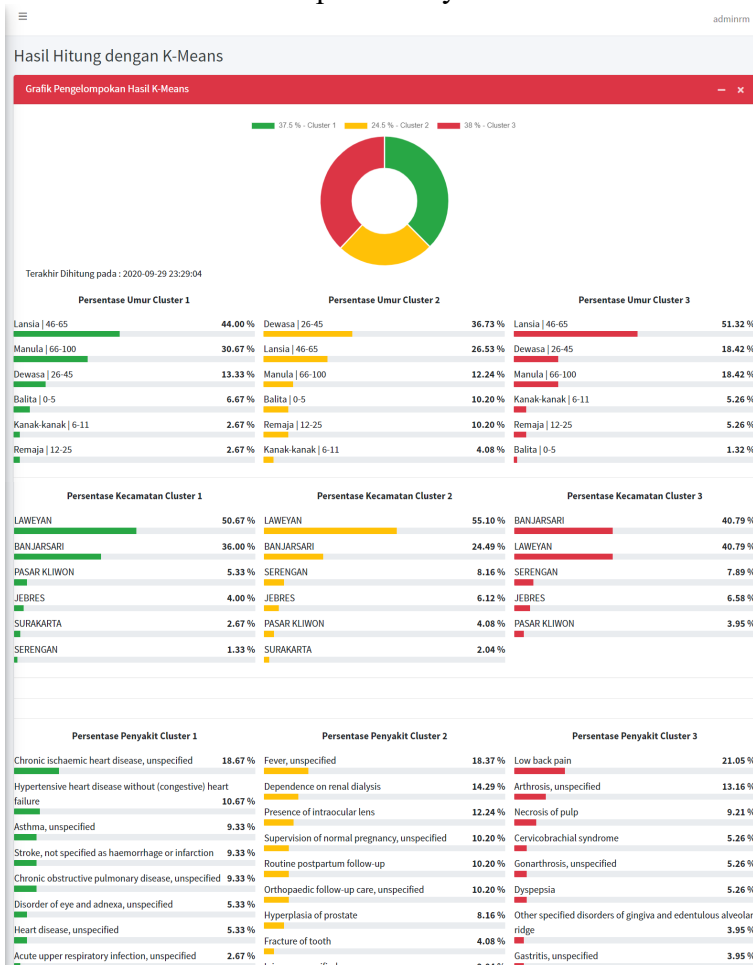
Data Ke-	Umur (Tahun)	ICD (Kode ICD 10)	Kecamatan (ID Kecamatan)	Cluster (Kelompok)
1	2	4	1	1
2	6	35	3	1
3	4	9	1	1
4	4	12	2	1
5	4	12	3	1

Gambar 3. Halaman Hasil Perhitungan

Adapun langkah-langkah algoritma pada metode clustering k-means adalah sebagai berikut:

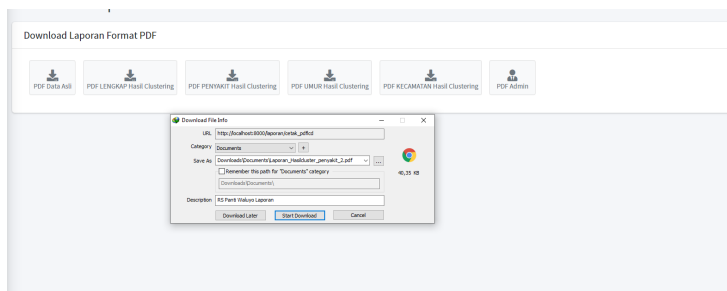
- 1) Tentukan nilai banyaknya cluster
- 2) Tentukan nilai centroid pusat
- 3) Hitung jarak euclidean data dengan masing-masing centroid
- 4) Hitung nilai jarak terkecil dan masukan data kedalam anggota cluster berdasarkan jarak yang terdekat

- 5) Hitung nilai centroid baru
 - 6) Hitung jarak euclidean data dengan masing-masing centroid baru
 - 7) Hitung nilai jarak terkecil dan masukan data kedalam anggota cluster berdasarkan jarak yang terdekat
 - 8) Berhenti apabila anggota cluster sudah tidak berubah
- Menampilkan grafik dan tabel hasil perhitungan K-means beserta rincian informasi setiap clusternya.



Gambar 4. Halaman grafik perhitungan

Menampilkan menu unduh data laporan dengan ekstensi file pdf kedalam penyimpanan komputer.



Gambar 5. Halaman unduh laporan



RUMAH SAKIT PANTI WALUYO SURAKARTA
Laporan Data Clustering Penyakit

Jl. Jend Ahmad Yani No.1 Surakarta, Jawa Tengah 57143 - Phone 0271-712077

Cluster 1 75 data 37.5 % dari total data 200		
Lansia	46-65 th	44.00 %
Manula	66-100 th	30.67 %
Dewasa	26-45 th	13.33 %
Balita	0-5 th	6.67 %
Kanak-kanak	6-11 th	2.67 %
Remaja	12-25 th	2.67 %
Cluster 2 49 data 24.5 % dari total data 200		
Dewasa	26-45 th	36.73 %
Lansia	46-65 th	26.53 %
Manula	66-100 th	12.24 %
Balita	0-5 th	10.20 %
Remaja	12-25 th	10.20 %
Kanak-kanak	6-11 th	4.08 %
Cluster 3 76 data 38 % dari total data 200		
Lansia	46-65 th	51.32 %
Dewasa	26-45 th	18.42 %
Manula	66-100 th	18.42 %
Kanak-kanak	6-11 th	5.26 %
Remaja	12-25 th	5.26 %
Balita	0-5 th	1.32 %
Cluster 1 75 data 37.5 % dari total data 200		
LAWEYAN		50.67 %
BANJARSARI		36.00 %
PASAR KLIWON		5.33 %
JEBRES		4.00 %
SURAKARTA		2.67 %
SERENGAN		1.33 %

Gambar 6. Hasil dokumen laporan

KESIMPULAN

Cluster 1 mempunyai anggota sejumlah 75 data atau 37,5 % dari total keseluruhan data. *Cluster 2* mempunyai anggota sejumlah 49 data atau 24,5 % dari total keseluruhan data. *Cluster 3* mempunyai anggota sejumlah 76 data atau 38 % dari total keseluruhan data. Penyakit terbanyak diderita pada *cluster1* yaitu: *Chronic ischaemic heart disease, unspecified* sebesar 18,67% dari total data *cluster 1*. Penyakit terbanyak diderita pada *cluster2* yaitu: *Fever, unspecified* sebesar 18,37% dari total data *cluster 2*. Penyakit terbanyak diderita pada *cluster3* yaitu: *Low back pain* sebesar 21,05% dari total data *cluster 3*. Rentang umur terbesar pada *cluster 1* berada pada kategori usia Lansia (46-65 th) sebesar 44 % dari keseluruhan data *cluster 1*. Rentang umur terbesar pada *cluster 2* berada pada kategori usia Lansia (46-65 th) sebesar 26,53 % dari keseluruhan data *cluster 2*. Rentang umur terbesar pada *cluster 3* berada pada kategori usia Lansia (46-65 th) sebesar 51,32 % dari keseluruhan data *cluster 3*. Wilayah dengan penderita terbanyak pada *cluster 1* yaitu Laweyan sebesar 50,67 % dari total data *cluster 1*. Wilayah dengan penderita terbanyak pada *cluster 2* yaitu Laweyan sebesar 55,10 % dari total data *cluster 2*. Wilayah dengan penderita terbanyak pada *cluster 3* yaitu Laweyan dan Banjarsari masing-masing sebesar 40,79 % dari total data *cluster 3*.

DAFTAR PUSTAKA

- Aditya, Mikael W. K. M. 2017. **Penerapan Metode K-Means Clustering Untuk Mengelompokkan Potensi Produksi Buah-Buahan di Provinsi Daerah Istimewa Yogyakarta**. Skripsi. Universitas Sanata Dharma Yogyakarta.
- Anisa, Yunita Wandyra. 2016. **Rekayasa perangkat lunak pengendalian inventori menggunakan metode sma (single moving average) berbasis ajax (asynchronous javascript and xml) (studi kasus: ptp nusantara vi (persero) unit usaha kayu aro)**. Jurnal. TEKNOIF. Vol 4 No 2. Institute Teknologi Padang.
- Ayu, Indah Septrianingrum, dkk. 2016. **Perancangan Dan Pengembangan Prototype Sistem Parkir**. Jurnal. KLIK. Vol 3 No 3. Universitas Lambung Mangkurat.

- Bastian, Adi dkk. 2018. **Penerapan Algoritma K-Menas Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)**. Jurnal Sistem Informasi. Vol 14 No 1. Universitas Majalengka.
- Chafid, Nurul dan Ismail Ari Wibowo. 2018. **Implementasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Demam Berdarah Di Kota Tangerang Selatan Menggunakan Algoritma K-Means**. Jurnal. Satya Inforamtika. Vol 3 No 1. Tangerang.
- Eliana dan Sri Sumiati. 2016. **Kesehatan Masyarakat**. E-book. Pusdik SDM Kesehatan. Desember 2016. Jakarta.
- Fajarianto, Otto. 2016. **Prototype Pelayanan Akademik Terhadap Komplain Mahasiswa Berbasis Mobile**. Jurnal. LENTERA ICT. Vol 3 No 1. STMIK Bina Sarana Global.
- Hadijah. 2016. **Analisis Kualitas Pelayanan Rawat Inap Di Rumah Sakit Umum Daerah Undata Palu Provinsi Sulawesi Tengah**. Jurnal. Katalogis. Vol 4 No7. Universitas Tadulako.
- Hakam, Fahmi, dkk. 2017. **Analisis Sistem Dan Teknologi Informasi Sebagai Acuan Dalam Perancangan Rencana Strategis Sistem Informasi Dan Teknologi Informasi (RENSTRA SI/TI) Di Rumah Sakit Islam Yogyakarta PDHI**. Jurnal. JSI. Vol 9 No 1. Universitas Surabaya.
- Hendini, Ade. 2016. **Pemodelan UML Sistem Informasi Monitoring Penjualan Dan Stok Barang (Studi Kasus: Distro Zhezha Pontianak)**. Jurnal. KHATULISTIWA. Vol 4 No 2. AMIK BSI Pontianak.
- Ika, Aulia Pratiwi, dkk. 2019. **Analisis Kepuasan Pasien Terhadap Kualitas Layanan Poli Gigi di Rumah Sakit Banyu Manik Semarang**. Jurnal. Manajemen Kesehatan Indoensia. Vol 7 No 1. Universitas Diponegoro.
- Irwan, SKM. 2017. **Epidemiologi Penyakit Menular**. CV Absolute Media. Yogyakarta
- Khrisna, Anindya Wardhani. 2016. **Implementasi Algoritma K-Means untuk pengelompokkan penyakit pasien pada puskesmas kajen pekalongan**. Jurnal. TRANSFORMATIKA. Vol 14 No 1. Universitas Diponegoro.
- Normala, Evi. 2018. **Implementasi K-Means Clustering Untuk Pengelompokan Penderita Penyakit Berdasarkan Analisis**

- Asociation Rule. SKRIPSI.** Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Lolong, Steven dkk. 2016. **Aplikasi Kamus Penyakit Manusia Berbasis Android.** Jurnal. COGITO. Vol 2 No 2. Universitas Klabat.
- Prasetijono, Purwito Soengeng. 2009. **Perancangan Sistem Informasi Pemanfaatan Kamar Operasi (OK) Rumah Sakit Islam Sultan Agung Semarang.** *Tesis.* Universitas Diponegoro Semarang.
- Rianto,Aurelia. 2017. **Implementasi Metode K-Means Clustering Dalam Mengelompokkan Emosi Senang, Marah, dan Netral Berdasarkan Vokal Manusia.** Skripsi. Universitas Multimedia Nusantara.
- Rusdiansyah, 2018. **Membangun Prototype Sistem Informasi Arsip Elektronik Surat Perjanjian Kerjasama Pada Business Support Departement.** Jurnal. Pilar Nusantara Mandiri. Vol 14 No 2. Universitas Bina Sarana Informatika.
- Silitongga, Parasian D P dan Romanus Damanik. 2016. **Implementasi Algoritma K-menas Clustering pada analisis penyebaran penyakit pasien pengguna Badan Penyelenggara Jaminan Sosial (BPJS).** Jurnal. JTIUST. Vol 01 No 01. Unika Santo Thomas.
- Suendri. 2018. **“Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus : UIN Sumatera Utara Medan)”.** Jurnal. ALGORITMA. Vol 3 No 1. Universitas Islam Negeri Sumatera Utara.
- Vitalocca Dyah, dkk. 2018. **Perancangan Sistem Informasi Data Dosen Jurusan Pendidikan Teknik Elektro Universitas Negeri Makassar Berbasis Web.** Jurnal. MEKOM. Vol 5 No 1. Universitas Negeri Makassar.
- Yustiningsih, Rini 2019. **“10 Ancaman Kesehatan Terbesar Tahun 2019”** <https://www.solopos.com/10-ancaman-kesehatan-terbesar-tahun-2019-967164>. Diakses pada 20 Desember 2019.