

KLUSTERISASI PENYAKIT ENDEMIS PADA KECAMATAN SABU BARAT, KABUPATEN SABU RAIJUA MENGGUNAKAN ALGORITMA K-MEANS

Wenefrida T. Ina¹, Yefta Mesakh², Stephanie E. Pella³

^{1,2,3} Program Studi Teknik Elektro, Fakultas Sains dan Teknik, Universitas Nusa Cendana
Jln. Adisucipto – Penfui, Telp. 03080-881597, Fax. 0380-881557
Email: wenefrida_ina@staf.undana.ac.id, yeftaam@gmail.com, s.i.pella@gmail.com

Info Artikel

Histori Artikel:
Diterima Feb 28, 2022
Direvisi Mar 14, 2022
Disetujui Mar 29, 2022

ABSTRACT

Information technology can be applied to identify endemic diseases in an area, in this case Sabu Raijua Regency. Endemic diseases can be identified early using the Clustering K-Means method where this method partitions data into one or more clusters/groups, so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into groups. another group. The data used in this study are medical record data at the Seba Health Center as many as 1020 data with year, village, diagnosis, age and gender variables. Due to the large amount of data, the K-Means Clustering process will use Weka 8.5 as a tool.

The results of this study indicate the characteristics and patterns of endemic diseases in the service area of the Seba Health Center with variables of year, village, diagnosis, age and gender, the characteristics used are based on the most optimal number of clusters. The most optimal number of clusters can be found using the Elbow Method. The results of clustering of 1020 medical record data showed that the most optimal number of clusters was 2 clusters with the characteristics of ARI diagnosis.

Keywords: K-Means Clustering, Elbow Method, WEKA, Medical Records.

ABSTRAK

Teknologi informasi dapat diterapkan untuk mengidentifikasi penyakit endemis pada suatu wilayah dalam hal ini Kabupaten Sabu Raijua. Penyakit endemis dapat diidentifikasi dini dengan metode Clustering K-Means dimana metode ini mempartisi data ke dalam bentuk satu atau lebih cluster/kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain. Data yang dipakai dalam penelitian ini yaitu data rekam medis pada Puskesmas Seba sejumlah 1020 data dengan variabel tahun, desa, diagnosa, usia dan gender. Dikarenakan data yang banyak, proses Clustering K-Means akan menggunakan Weka 8.5 sebagai alat bantu.

Hasil dari penelitian ini menunjukkan karakteristik dan pola penyakit endemi pada daerah pelayanan Puskesmas Seba dengan variabel tahun, desa, diagnosa, usia dan gender, karakteristik yang dipakai berdasarkan jumlah klaster yang paling optimal. Jumlah klaster paling optimal dapat dicari menggunakan Metode Elbow. Hasil klasterisasi dari 1020 data rekam medis menunjukkan jumlah klaster yang paling optimal adalah 2 klaster dengan karakteristik diagnose ISPA.

Kata Kunci: Klasterisasi K-Means, Metode Elbow, WEKA, Rekam Medis

Penulis Korespondensi:

Wenefrida Tulit Ina, ST., MT
Program Studi Teknik Elektro Fakultas Sains dan Teknik,
Universitas Nusa Cendana,
Jl. Adisucipto Penfui - Kupang.
Email: wenefrida_ina@staf.undana.ac.id



1. PENDAHULUAN

Rekam medis merupakan salah satu bukti rekam jejak penyakit pasien yang berisi data pribadi pasien tanggal pengobatan, diagnose penyakit pasien, jenis obat yang diberikan serta riwayat sakit pasien. Dengan adanya bukti tertulis tersebut maka data rekam medis yang diberikan dapat dipertanggungjawabkan dengan tujuan sebagai penunjang tertib administrasi dalam upaya peningkatan pelayanan kesehatan masyarakat.

Data rekam medis yang terdapat pada fasilitas kesehatan selalu bertambah setiap hari seiring dengan jumlah kunjungan pasien. Namun, data yang tersimpan hanya sebatas memberikan grafik atau statistik jumlah pasien yang berobat dengan penyakit yang dideritanya beserta laporan kepulauan pasien tersebut. Belum ada pengolahan data rekam medis yang merujuk pada penggalian informasi penting yang menghasilkan suatu pola yang dapat digunakan oleh dinas kesehatan untuk membuat kebijakan-kebijakan untuk meningkatkan pelayanan kesehatan kepada masyarakat. Mengenai pola dari kecenderungan penyakit yang diderita oleh masyarakat masih belum digali untuk dijadikan acuan apabila melakukan pencegahan penyakit [1]. Menurut [2] Data rekam medis dapat diklasifikasi untuk mengetahui pola penyebaran penyakit pada masyarakat. Data rekam medis ini pun dapat digunakan untuk mengetahui wilayah-wilayah penyebaran penyakit endemis.

Dalam hal ini pengolahan data rekam medis dapat diterapkan untuk mengidentifikasi penyakit endemis pada Kabupaten Sabu Raijua. Penyakit endemis dapat diidentifikasi dini dengan metode K-Means berdasarkan penelitian yang telah dilakukan sebelumnya yakni Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia [4] Pengelompokan Data Rekam Medis untuk Mengetahui Penyakit Endemi di suatu Daerah Menggunakan K-Means Clustering [1] dan Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance [3]

Berdasarkan data diatas dan penelitian sebelumnya maka penulis akan melakukan penelitian dengan judul “Klasterisasi Penyakit Endemis Studi Kasus Kecamatan Sabu Barat, Kabupaten Sabu Raijua Menggunakan Algoritma K-Means” yang diharapkan dapat

membantu pemerintah dalam melihat penyakit endemis yang ada pada Kecamatan Sabu Barat.

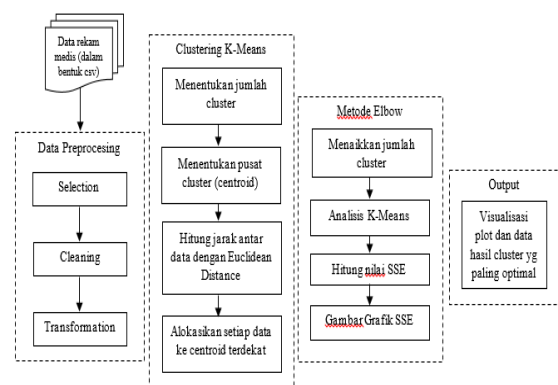
2. METODE PENELITIAN

Secara umum alur pada penelitian ini akan dimulai dari pengambilan atau pengumpulan data rekam medis, kemudian data tersebut akan melalui proses *selection*, *cleaning* dan *transformation*, ketiga langkah tersebut bertujuan agar menghasilkan data yang sesuai dengan data yang diperlukan baik dari sisi variabel, nilai inisialisasi, format dan data tidak memiliki kesalahan penulisan.

Kemudian pada proses *data mining* yaitu *clustering K-Means* adapun langkah-langkah klasterisasi dengan menentukan jumlah *cluster* yang diinginkan, setelah itu menentukan pusat *cluster (centroid)* dari setiap variabel di setiap *cluster*, setelah itu menghitung jarak data ke setiap *centroid* dengan persamaan *Euclidean Distance*, setelah mendapatkan nilai jarak data ke setiap *centroid* langkah selanjutnya adalah mengalokasikan setiap data ke *centroid* terdekat untuk menentukan *cluster* dari setiap data yang ada.

Selanjutnya menentukan jumlah klaster yang paling optimal menggunakan Metode Elbow. Beberapa jumlah klaster akan dipilih kemudian dilakukan analisis K-Means untuk setiap klaster kemudian menghitung nilai SSE, nilai-nilai tersebut akan dibuat dalam sebuah grafik, dengan prinsip Metode Elbow dapat dilihat jumlah klaster mana yang paling optimal.

Langkah terakhir adalah visualisasi data dengan plot dan hasil *cluster* untuk melihat pola penyakit endemis di daerah pelayanan Puskesmas Seba sesuai dengan karakteristik klaster yang paling optimal. Adapun arsitektur umum setiap tahapan metode yang digunakan dalam penelitian ini ditunjukkan pada Gambar dibawah.



Gambar 2.1 Arsitektur Umum

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan *K-Means Clustering* untuk mengelompokkan data-data rekam medis yang diambil pada Puskesmas Seba di Kecamatan Sabu Barat selama 4 tahun terakhir. Data-data tersebut direkap dari setiap form rekam medis dan di-*input*-kan ke MS.Excel. Terdapat beberapa tahapan yang dilakukan untuk memulai proses *clustering* yakni *Data Selection*, *Data Cleaning* dan *Data Transformation* yang dilakukan pada aplikasi MS.Excel kemudian tahap *Data Mining* dan *Evaluation*.

Pada penelitian juga dilakukan perhitungan Analisis K-Means dan Analisis Metode Elbow secara Manual dengan mengambil 50 sampel data setelah itu dibandingkan dengan hasil analisis yang dilakukan pada aplikasi WEKA 8.5. Hasil akhir yang diperoleh adalah terjadinya perbedaan jumlah iterasi dan nilai SSE dikarenakan terjadinya perbedaan perhitungan berkaitan dengan nilai angka dibelakang koma, untuk 50 sampel data proses perhitungan manual terjadi sangat panjang sehingga untuk perhitungan 1020 data membutuhkan Tool Data Mining dalam hal ini Aplikasi WEKA 8.5.

Tahap *clustering* akan dilakukan pada aplikasi WEKA 8.5 sebagai *tools data mining* untuk membantu dikarenakan data yang sangat banyak sehingga sangat sulit untuk dilakukan secara manual. Metode Elbow akan digunakan untuk menentukan jumlah *cluster* yang paling optimal. Hasil akhir dari penelitian ini yaitu dapat menunjukkan pola penyakit endemi pada daerah pelayanan Puskesmas Seba berdasarkan karakteristik dari jumlah *cluster* yang paling optimal.

3.1. Data Selection

Pada tahap ini data dilakukan pengumpulan data dari form rekam medis yang ada pada Puskesmas Seba, kemudian pemilihan data yang sesuai dengan variable/parameter yang digunakan dalam penelitian ini yakni data waktu, desa, diagnosa penyakit, nama pasien dan umur. Pada tahap ini data yang dipakai adalah data tahun 2018 sampai tahun 2021, jadi beberapa diagnosa penyakit sebelum tahun 2018 tidak dipakai atau dipilih.

3.2. Data Cleaning

Pada tahap ini dilakukan pengambilan data-data yang sesuai dengan variabel yang akan digunakan, tahap ini dilakukan proses pemisahan beberapa data seperti data diagnosa penyakit yang terdapat pada pasien dan waktu yang sama agar data tersebut dapat diinisialisasi, data waktu yang

sebelumnya dicantumkan bulan pada tahap ini hanya menyisahkan data tahun saja sesuai variabel yang dipakai dalam penelitian ini yakni tahun 2018 (18), 2019 (19), 2020 (20) dan 2021 (21), Tahap ini juga dilakukan pengelompokan kategori usia menurut WHO yakni Balita (<5), Kanak-kanak (5 s/d 11), Remaja (12 s/d 25), Dewasa (26 s/d 45), Lansia (46 s/d 65) dan Manula (>65), kemudian dilakukan perbaikan terhadap kesalahan penulisan pada nama diagnosa penyakit, kategori usia dan kesalahan penulisan lain yang terdapat dalam data.

Setelah melalui proses *cleaning* data yang sebelumnya sejumlah 835 menjadi 1020 data yang akan dilakukan proses *clustering* untuk mendapatkan pola penyakit endemis pada daerah pelayanan Puskesmas Seba.

3.3. Data Transformation

Pada 1020 data diatas terdapat data non-numerik seperti data desa, diagnosa, kategori umur dan gender, agar data tersebut dapat diolah menggunakan algoritma *k-means* maka data-data non-numerik perlu diinisialisasi ke data numerik, data-data tersebut diubah secara manual di excel.

Tabel 3.1. Data Rekam Medis setelah proses inisialisasi

| No | Tahun | Desa | Diagnosa | Kategori | Gender |
|------|-------|-------|----------|----------|--------|
| 1 | 2 | 9 | 18 | 2 | 1 |
| 2 | 2 | 9 | 10 | 2 | 1 |
| 3 | 2 | 9 | 1 | 2 | 1 |
| 4 | 3 | 9 | 1 | 2 | 1 |
| 5 | 1 | 9 | 1 | 5 | 2 |
| 6 | 2 | 9 | 8 | 5 | 2 |
| | | | | | |
| ... | | | | | |

3.4. Analisis Clustering K-Means

Pada tahap ini akan dilakukan analisis *clustering* menggunakan metode *K-Means*. Tahapan-tahapan yang akan dilakukan adalah sebagai berikut:

3.4.1. Menentukan Jumlah Cluster

Tahapan awal dalam proses *clustering* adalah menentukan jumlah *cluster* k. pada tahap ini k akan di tentukan dari k terkecil sebagai awal dari analisis ini, kemudian jumlah *cluster* k akan dinaikkan untuk mencari jumlah *cluster* k yang paling optimal menggunakan Metode Elbow.

3.4.2. Menentukan Pusat Cluster (Centroid)

Tahap selanjutnya adalah menentukan nilai pusat *cluster* (*centroid*) awal secara acak dari data yang

telah diinputkan. Tentukan *centroid* dari setiap variabel yang dipakai yaitu nilai *centroid* tahun, nilai *centroid* desa, nilai *centroid* diagnosa, nilai *centroid* kategori umur dan nilai *centroid* gender.

3.4.3. Menghitung Jarak Data dengan Euclidean Distance

Perhitungan *euclidean distance* dapat dengan persamaan 3.1:

$$d(xi, \mu_j) = \sqrt{(xia - \mu_ja)^2 + (xib - \mu_jb)^2} \dots\dots (3.1)$$

Dengan ketentuan sebagai berikut:

xi : data variabel

μ_j : *centroid* pada *cluster* ke-j

Perhitungan *Euclidean distance* bertujuan untuk mendapatkan jarak dari masing-masing data ke pusat data (*centroid*).

3.4.4. Alokasikan Data pada Cluster Terdekat

Proses selanjutnya yakni mengalokasikan setiap data ke suatu *cluster* berdasarkan jarak terdekat dari *centroid*-nya. Jika data pertama diperoleh jarak terdekat dengan *centroid* pertama (C_0), maka data tersebut akan menjadi anggota *cluster* pertama, dan seterusnya sampai semua data selesai.

3.4.5. Hitung Nilai Centroid Baru

Jika persebaran data masih banyak yang berada diluar *cluster* maka dilakukan perhitungan nilai *centroid* dengan cara menghitung nilai rata-rata dari masing-masing parameter/variabel setiap data yang menjadi anggota *cluster* tertentu. Untuk menghitung nilai *centroid* baru digunakan persamaan 3.2.:

$$\mu = \frac{\sum_{i=1}^n xi}{n} \dots\dots (3.2)$$

Dengan keterangan sebagai berikut:

μ = *centroid* pada *cluster*

xi = data ke-*i*

n = banyaknya data/jumlah data yang menjadi anggota *cluster*

selanjutnya akan dilakukan perhitungan iterasi kedua dengan menghitung jarak setiap data terhadap nilai *centroid* yang baru sesuai persamaan *Euclidean distance* (3.1), proses ini akan terus berlanjut sampai pada kedudukan anggota setiap *cluster* tidak lagi berubah.

3.5. Analisis Metode Elbow

Pada tahap ini akan dilakukan analisis Metode Elbow yaitu dengan menghitung nilai *Sum of Squared Error* (SSE) dari setiap *cluster* yang ada seperti persamaan (2.3). Setelah data di klusterisasi dengan *k* terkecil selanjutnya nilai *k* akan dinaikkan sampai *k* tertentu yang kemudian akan diuji

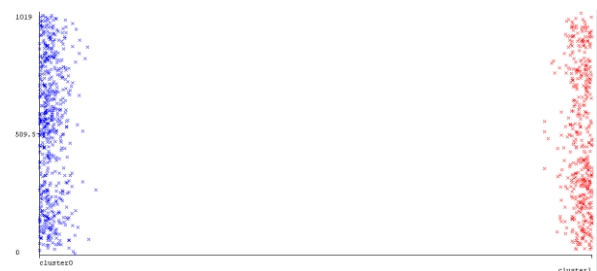
menggunakan Metode Elbow yaitu menghitung nilai SSE setiap *cluster* untuk mendapatkan nilai *cluster* *k* yang paling optimal, perhitungan pada algoritma *k-means* sangat menentukan hasil dari metode elbow sehingga memerlukan ketelitian saat melakukan klusterisasi, yang perlu diperhatikan bahwa nilai *centroid* tidak boleh berubah, sehingga perlu ditentukan terlebih dahulu semua nilai *centroid* [6]

3.6. Implementasi K-Means Clustering pada WEKA

Pada tahap ini proses *Clustering K-Means* dilakukan pada aplikasi WEKA untuk menemukan pola penyakit endemis pada daerah pelayanan Puskesmas Seba dikarenakan data yang terlalu banyak, penentuan jumlah *cluster* *k* juga dilihat pada hasil nilai SSE yang ditampilkan pada WEKA, WEKA juga membantu untuk memvisualisasikan penyebaran data-data di setiap *cluster* sehingga dapat dilihat bahwa *cluster* tertentu memiliki karakteristik tertentu. Data yang digunakan pada WEKA adalah data-data rekam medis yang telah melalui proses *selection*, *cleaning*, dan *transformation*. Agar data-data tersebut dapat diproses dalam aplikasi WEKA, data-data yang sebelumnya berupa Excel akan diubah ke format *Comma Separated Values* (.csv). Terdapat beberapa tahapan dalam aplikasi WEKA yakni proses *input data*, proses *filtering*, proses *K-Means Clustering* dan Pengujian Algoritma.

3.7. Hasil Clustering

Dari hasil pengolahan 1020 data rekam medis menggunakan WEKA dapat dilihat persentase penyebaran data-data yang digunakan ke setiap klaster yaitu 57% data pada *cluster 0* dan 43% data pada *cluster 1*. Untuk penyebaran setiap data ke setiap *cluster* dapat dilihat pada gambar 4.10. warna biru untuk C_0 , warna merah untuk C_1 , warna hijau untuk C_2 dan warna aqua untuk C_3 dengan keterangan sumbu X menunjukkan setiap *cluster* dan sumbu Y menunjukkan jumlah data.



Gambar 3.1. Plot Data Rekam Medis

Dari proses *clustering* didapatkan informasi sebagai berikut:

1. *Cluster* pertama (C_0) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia dewasa, dengan gender laki-laki, di desa Raeloro pada Tahun 2020.
2. *Cluster* kedua (C_1) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Nadawawi pada Tahun 2019.

3.8. Pengujian Sistem

Pengujian sistem dilakukan dengan menaikkan jumlah kluster pada k tertentu, pada penelitian ini nilai kluster dinaikkan sampai $k=16$ sehingga ketika nilai SSE didapatkan dan dibuat dalam sebuah grafik maka perubahan nilai setiap jumlah kluster lebih teliti. Berikut adalah beberapa karakteristik dari beberapa jumlah kluster dengan menggunakan 1020 data rekam medis. Pada pengujian sistem, proses *clustering* menggunakan dua *cluster* menghasilkan kesimpulan sebagai berikut.

1. *Cluster* pertama (C_0) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia dewasa, dengan gender laki-laki, di desa Raeloro pada Tahun 2020.
2. *Cluster* kedua (C_1) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Nadawawi pada Tahun 2019.

Proses *clustering* menggunakan tiga *cluster* menghasilkan kesimpulan sebagai berikut:

1. *Cluster* pertama (C_0) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia remaja, dengan gender perempuan, di desa Raeloro pada Tahun 2020.
2. *Cluster* kedua (C_1) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Raeloro pada Tahun 2019.
3. *Cluster* ketiga (C_2) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia balita, dengan gender laki-laki, di desa Ledekapaka pada Tahun 2018.

Proses *clustering* menggunakan lima *cluster* menghasilkan kesimpulan sebagai berikut:

1. *Cluster* pertama (C_0) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia remaja, dengan gender perempuan, di desa Raeloro pada Tahun 2020.
2. *Cluster* kedua (C_1) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang

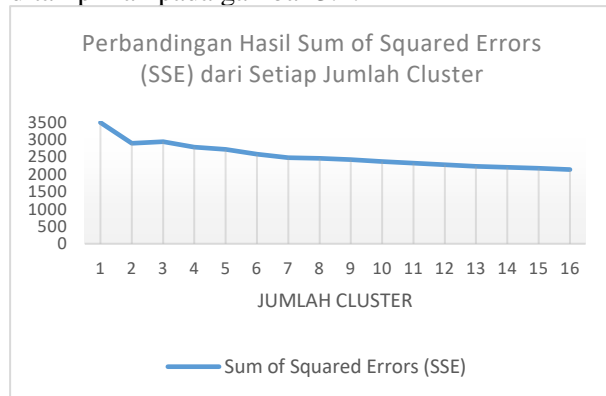
- kategori usia lansia, dengan gender perempuan, di desa Raeloro pada Tahun 2019.
3. *Cluster* ketiga (C_2) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia balita, dengan gender laki-laki, di desa Ledekapaka pada Tahun 2018.
4. *Cluster* keempat (C_3) memiliki karakteristik diagnosa penyakit DYSPEPSI, banyak menyerang kategori usia dewasa, dengan gender laki-laki, di desa Titinalede pada Tahun 2018.
5. *Cluster* kelima (C_4) memiliki karakteristik diagnosa penyakit HIPERTENSI, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Ledeanana pada Tahun 2019.

Proses *clustering* menggunakan tujuh *cluster* menghasilkan kesimpulan sebagai berikut:

1. *Cluster* pertama (C_0) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia remaja, dengan gender perempuan, di desa Raeloro pada Tahun 2020.
2. *Cluster* kedua (C_1) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Raeloro pada Tahun 2019.
3. *Cluster* ketiga (C_2) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia balita, dengan gender laki-laki, di desa Ledekapaka pada Tahun 2018.
4. *Cluster* keempat (C_3) memiliki karakteristik diagnosa penyakit DYSPEPSI, banyak menyerang kategori usia dewasa, dengan gender laki-laki, di desa Titinalede pada Tahun 2018.
5. *Cluster* kelima (C_4) memiliki karakteristik diagnosa penyakit HIPERTENSI, banyak menyerang kategori usia lansia, dengan gender perempuan, di desa Ledeanana pada Tahun 2019.
6. *Cluster* keenam (C_5) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia dewasa, dengan gender laki-laki, di desa Raemadia pada Tahun 2020.
7. *Cluster* ketujuh (C_6) memiliki karakteristik diagnosa penyakit ISPA, banyak menyerang kategori usia remaja, dengan gender perempuan, di desa Raenyale pada Tahun 2019.

Setelah pemrosesan *clustering* selesai, akan dilakukan pengujian dengan nilai $k=1$ sampai $k=16$ dengan menggunakan *Sum of Squared Errors* (SSE), nilai SSE dapat dilihat pada hasil *clustering* yang ada pada Aplikasi WEKA. Pengujian tersebut bertujuan untuk mengetahui performa *K-Means* dan menguji validitas *cluster*. Metode *Sum of Squared Errors* (SSE) memberikan informasi *error* jarak data ke

centroid. Jika semakin kecil nilai SSE maka semakin bagus hasil *clustering* yang diperoleh. Hasil perhitungan *Sum of Squared Errors* ditampilkan pada gambar 3.2.



Gambar 3.2. Grafik Hasil SSE dari Setiap Jumlah Cluster

Dari gambar 3.2., diketahui bahwa hasil *Sum of Squared Errors* (SSE) yang membentuk siku ada pada $k=1$ ke $k=2$ setelah itu nilai SSE mulai konstan, sehingga berdasarkan prinsip dari Metode Elbow dimana mencari dengan melihat grafik dari nilai SSE yang membentuk siku, serta dengan tujuan mencari nilai SSE terendah dengan jumlah *cluster* k terkecil dapat disimpulkan bahwa jumlah *cluster* k yang paling optimal berada pada jumlah *cluster* $k=2$ yaitu sebesar 2890.

4. KESIMPULAN

Dari hasil penerapan Algoritma *Clustering K-Means* pada 1020 data rekam medis dari tahun 2018 sampai awal Tahun 2021 pada daerah pelayanan Puskesmas Seba dapat diambil kesimpulan sebagai berikut.

1. Hasil klasterisasi dari 1020 data menggunakan Algoritma K-Means menunjukkan bahwa Algoritma K-Means dapat diterapkan pada kumpulan data rekam medis untuk menghasilkan pola penyakit endemis pada daerah pelayanan Puskesmas Seba.
2. Metode Elbow dapat digunakan untuk mencari jumlah *cluster* k yang paling optimal, yaitu dengan melihat model grafik yang dibentuk dari nilai SSE dari setiap *cluster*, pada penelitian ini jumlah *cluster* yang paling optimal adalah $k=2$.
3. Dari data-data yang dikumpulkan dan diklasterisasi menggunakan Algoritma K-Means dengan karakteristik *cluster* yang paling optimal berdasarkan Metode Elbow maka dapat dilihat bahwa dari kedua *cluster* tersebut memiliki karakteristik diagnosa yang sama yaitu ISPA.

DAFTAR PUSTAKA

- [1] Lisa Olivia., Pengelompokan Data Rekam Medis untuk Mengetahui Penyakit Endemi di Suatu Daerah Menggunakan K-Means Clustering, Medan: Skripsi Universitas Sumatera Utara, 2019.
- [2] Wenefrida Tulit Ina, "Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5" Jurnal Media Elektro, Vol 1 Edisi 3, April 2013.
- [3] M Hariyanto, R T Shita, "Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance," skripsi Universitas Budi Luhur, 2018.
- [4] A Bastian, H Sujadi, G Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)" Jurnal Sistem Informasi, Jil.14, No.1, 2018.
- [5] Elly Muningsih, Sri Kriswati, "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan" Jurnal Joutica, vol.3, No.1, 2018.
- [6] NPE Merliana, Ernawati, Alb. Joko Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering", Prosiding Sendi_U, 2016. ISBN : 978-979-3649-81-8.
- [7] Anita F., Antonito H. Cabral, Gangga Anuraga, "K-Means Clustering Dengan Metode Elbow Untuk Pengelompokan Kabupaten Dan Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan", Prosiding SNHRP-1/2018, 21 Desember 2018, ISBN : 978-602-5793-40-0.
- [8] Hasyrif SY, Rismayani, Asrul Syam,"Data Mining Menggunakan algoritma K-Means pengelompokan Penyebaran Diare di Kota Makassar", Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi, Vol. VIII, No.1, Hal : 73-82, Februari 2019
- [9] Jeri Wardana, Sarjon Defit, Sumijan,"Klarerisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means" Jurnal Informasi dan Teknologi, Vo.2, No.4. hal:119-125. ISSN: 2714-9730 (*electronic*), 2020.
- [10] Nurul Chafid, Ismail Ari Wibowo," Implementasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Demam Berdarah di Kota Tangerang Selatan Menggunakan Algoritma K-Means (Studi Kasus: Dinas Kesehatan Tangerang Selatan)", Jurnal Satya Informatika Vol.3 No.1, hal: 11-24, Mei 2018