

SKRIPSI

**OPTIMASI K-MEANS CLUSTERING PADA PENYAKIT
MENULAR DENGAN ALGORITMA PARTICLE SWARM
OPTIMIZATION**

Disusun dan diajukan oleh:

**ANDI RUSMIATI
D121 19 1079**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023**

LEMBAR PENGESAHAN SKRIPSI

OPTIMASI K-MEANS CLUSTERING PADA PENYAKIT MENULAR DENGAN ALGORITMA PARTICLE SWARM OPTIMIZATION

Disusun dan diajukan oleh

ANDI RUSMIATI
D121191079

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian
Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 9 Juni 2023
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

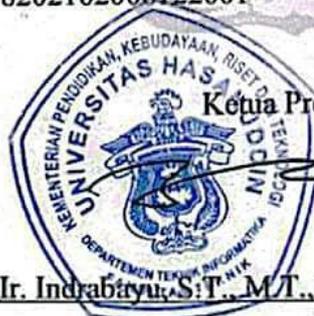
Pembimbing Pendamping,



Elly Warni, S.T., M.T.
NIP 198202162008122001

Ir. Christoforus Yohanes, M.T.
NIP 196007161987021002

Ketua Program Studi,



Prof. Dr. Ir. Indrabayuni, S.T., M.T., M.Bus.Sys., IPM., ASEAN.Eng.

NIP 197507162002121004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini;

Nama : Andi Rusmiati
NIM : D121191079
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Optimasi K-Means Clustering Pada Penyakit Menular Dengan Algoritma Particle Swarm Optimization

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 11 Juni 2023

Yang Menyatakan



Andi Rusmiati

ABSTRAK

ANDI RUSMIATI. *Optimasi K-means Clustering Pada Penyakit Menular Dengan Algoritma Particle Swarm Optimization* (dibimbing oleh Elly Warni dan Christoforus Yohannes)

Penyakit menular yang tingkat penularannya cukup tinggi yang terjadi di Puskesmas Bontomarannu, yakni tuberkulosis, DBD, dan demam tifoid. Hal ini menyulitkan pihak puskesmas dalam mengelompokkan faktor apa yang menjadi penyebab bertambahnya penyakit tersebut setiap harinya karena salah satu alasan yakni keterbatasan tenaga medis untuk mendeteksi secara manual indikator penyebab munculnya penyakit tersebut melalui data rekam medis. Berdasarkan data yang diperoleh melalui kuesioner yang diberikan kepada pasien puskesmas bontomarannu, terdapat beberapa variable yang berhubungan dengan penyakit menular. Penelitian ini bertujuan untuk membuat pengelompokan data penyakit menular (tuberkulosis DBD, dan tifoid) per desa di Kabupaten Gowa. Metode yang digunakan adalah *K-means clustering* yang akan dioptimalkan menggunakan algoritma PSO sehingga diperoleh hasil yang lebih baik. Hasil *clustering* yang terbentuk yaitu sebanyak 52 *cluster*. Dimana tiap desa dapat termasuk dalam 1 *cluster* atau lebih yang menunjukkan *cluster* yang paling berpengaruh terhadap kasus penyakit menular di desa tersebut. Hasil dari penelitian ini diperoleh hasil bahwa algoritma PSO dapat mengoptimalkan *K-means clustering* dengan diperoleh nilai SSE, *Silhouette coefficient*, dan *Davies Bouldin Index* (DBI) yang lebih kecil dibanding hanya menggunakan *K-means*. Hasil *clustering* ditampilkan dalam bentuk visualisasi WebGIS yang menampilkan informasi *cluster* tiap desa di Kabupaten Gowa.

Kata kunci: *tuberculosis, dbd, tifoid, clustering, K-means clustering, pso, DBI, webGIS*

ABSTRACT

ANDI RUSMIATI. *Optimization of K-means Clustering in Infectious Diseases Using the Particle Swarm Optimization Algorithm* (supervised by Elly Warni and Christoforus Yohannes)

Infectious diseases with a high transmission rate are occurring at the Bontomarannu Health Center, namely tuberculosis, DBD and typhoid fever. This makes it difficult for the puskesmas to classify what factors are the cause of the increase in the disease every day for one reason, namely the limitations of medical personnel to detect manually the indicators that cause the emergence of the disease through medical record data. Based on data obtained through direct interviews with patients at the Bontomarannu Public Health Center, there are several variables related to infectious diseases. This study aims to classify data on infectious diseases (tuberculosis, DBD, and typhoid) per village in Gowa Regency. The method used is *K-means clustering* which will be optimized using the PSO algorithm so that better results are obtained. The *clustering* results formed are as many as 52 *clusters*. Where each village can be included in 1 *cluster* or more which shows the cluster that has the most influence on cases of infectious diseases in the sub-district. The results of this study show that the PSO algorithm can optimize K-means *clustering* by obtaining smaller SSE, *Silhouette coefficient*, and *Davies Bouldin Index* (DBI) values than using only *K-means Clustering* results are displayed in the form of a WebGIS visualization that displays cluster information for each village in Gowa Regency.

Keywords: *tuberculosis, dbd, typhoid, clustering, K-means clustering, pso, DBI, webGIS*

DAFTAR ISI

ABSTRAK	i
ABSTRACT	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	x
KATA PENGANTAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Penyakit Menular	6
2.2 Tuberkulosis	6
2.3 Demam Tifoid	8
2.4 Demam Berdarah Dengue	9
2.5 Data Mining	11
2.6 <i>K-means Clustering</i>	12
2.7 <i>Particle Swarm Optimization</i>	15
2.8 <i>Sum of Squared Error (SSE)</i>	16
2.9 <i>Silhouette Coefficient</i>	17
2.10 <i>Davies-Bouldin Index</i>	18
2.11 Web GIS	19
BAB III METODOLOGI PENELITIAN	21
3.1 Waktu dan Lokasi Penelitian	21
3.2 Instrumen Penelitian	21
3.3 Tahapan Penelitian	22
3.4 Teknik Pengambilan Data	24
3.5 Perancangan Sistem	24
3.6 <i>Clustering</i> Algoritma <i>K-means</i>	29
3.7 Optimasi <i>K-means</i> dengan PSO	29
3.8 Evaluasi <i>Clustering</i>	30
3.9 Pattern Evaluation	30
3.10 Visualisasi WebGIS	30
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Hasil Penelitian	31
4.1.1 <i>Clustering</i> berdasarkan Jumlah TB, Umur, dan Kondisi Lingkungan	31
4.1.2 <i>Clustering</i> berdasarkan Jumlah DBD, Umur, dan Kondisi Lingkungan	49
4.1.3 <i>Clustering</i> berdasarkan Jumlah Tifoid, umur, dan kondisi lingkungan	60

4.1.4	<i>Clustering</i> berdasarkan Jumlah TB, Umur, dan Kebiasaan Pola Hidup .	72
4.1.5	<i>Clustering</i> berdasarkan Jumlah DBD, Umur, dan Kebiasaan Pola Hidup	78
4.1.6	<i>Clustering</i> berdasarkan Jumlah Tifoid, Umur, dan Kebiasaan Pola Hidup	84
4.2	Pembahasan	90
4.2.1	Evaluasi Hasil <i>Cluster</i>	90
4.2.2	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah TB, Umur, dan Kondisi Lingkungan	92
4.2.3	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah DBD, Umur, dan Kondisi Lingkungan	94
4.2.4	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah Tifoid, Umur, dan Kondisi Lingkungan	96
4.2.5	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah TB, Umur, dan Kebiasaan Pola Hidup	98
4.2.6	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah DBD, Umur, dan Kebiasaan Pola Hidup	100
4.2.7	Interpretasi Hasil <i>Clustering</i> berdasarkan Jumlah Tifoid, Umur, dan Kebiasaan Pola Hidup	101
4.2.8	Observasi Hasil <i>Clustering</i>	102
4.2.9	Interpretasi Hasil <i>Clustering</i> untuk Setiap Desa di Kabupaten Gowa	104
4.2.10	Visualisasi WebGIS	136
BAB V KESIMPULAN DAN SARAN		143
5.1	Kesimpulan	143
5.2	Saran	144
DAFTAR PUSTAKA		146
LAMPIRAN		149

DAFTAR GAMBAR

Gambar 2. 1 Tahap data mining penemuan KDD (Han et al., 2011).....	11
Gambar 2. 2 Flowchart K-means clustering (Parteek Bhatia, 2019)	14
Gambar 2. 3 Arsitektur <i>WebGIS</i> (Fu, 2011)	20
Gambar 3.3.1 Tahapan Penelitian	22
Gambar 3.5.1 Sampel Data Penyakit Menular	26
Gambar 3.5.2 Sampel Feature Selection	27
Gambar 3.5.3 Sampel data sebelum dan setelah Multilable Binarizer	28
Gambar 3.5.4 Sampel data pasien per desa.....	28
Gambar 3.5.5 Sampel data setelah normalisasi.....	29
Gambar 4.1.1. 1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah tb, umur, dan kondisi lingkungan	32
Gambar 4.1.1.2 Inisialisasi jumlah cluster	33
Gambar 4.1.1.3 Centroid awal algoritma k-means	34
Gambar 4.1.1.4 Sampel jarak tiap data terhadap centroid pada iterasi pertama ...	35
Gambar 4.1.1.5 Sampel penentuan cluster 1 pada iterasi pertama.....	36
Gambar 4.1.1.6 Nilai centroid yang baru setelah iterasi pertama	36
Gambar 4.1.1.7 Nilai centroid yang dihasilkan tidak berubah pada iterasi ke-2 ..	37
Gambar 4.1.1.8 Sampel penentuan cluster pada iterasi ke-2	37
Gambar 4.1.1.9 Inisialisasi pusat cluster.....	38
Gambar 4.1.1.10 Perhitungan nilai SSE dan Silhouette.....	38
Gambar 4.1.1.11 Inisialisasi nilai V_0	38
Gambar 4.1.1.12 Update nilai velocity	39
Gambar 4.1.1.13 Update nilai centroid	39
Gambar 4.1.1. 14 Bandingkan nilai SSE baru dan lama.....	39
Gambar 4.1.1. 15 Perhitungan nilai DBI.....	40
Gambar 4.1.1.16 Perbandingan nilai centroid cluster berdasarkan jumlah TB, umur, dan kondisi lingkungan sebelum dan setelah optimasi PSO	41
Gambar 4.1.1.17 Sampel penentuan cluster berdasarkan jumlah TB, umur, dan kondisi lingkungan setelah optimasi	42
Gambar 4.1.1.18 Perbandingan Plot hasil cluster berdasarkan jumlah TB, umur, dan kondisi lingkungan sebelum dan sesudah optimasi.....	43
Gambar 4.1.1.19 Daftar desa pada cluster 1 berdasarkan jumlah TB, umur, dan kondisi lingkungan	47
Gambar 4.1.2. 1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah DBD, umur, dan kondisi lingkungan.....	50
Gambar 4.1.2.2. Perbandingan nilai centroid cluster berdasarkan Jumlah DBD, umur, dan kondisi lingkungan sebelum dan sesudah optimasi	52
Gambar 4.1.2.3 Sampel hasil cluster berdasarkan jumlah DBD, umur, dan kondisi lingkungan setelah optimasi	53
Gambar 4.1.2.4 Perbandingan plot hasil cluster berdasarkan jumlah DBD, umur, dan kondisi lingkungan sebelum dan sesudah optimasi.....	54
Gambar 4.1.2.5 Daftar desa pada cluster 1 berdasarkan jumlah DBD, umur, dan kondisi lingkungan.....	58

Gambar 4.1.3. 1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah tifoid, umur, dan kondisi lingkungan.....	61
Gambar 4.1.3.2 Perbandingan nilai centroid cluster berdasarkan jumlah tifoid, umur, dan kondisi lingkungan sebelum dan sesudah optimasi	64
Gambar 4.1.3. 3 Sampel hasil cluster berdasarkan jumlah tifoid, umur, dan kondisi lingkungan.....	65
Gambar 4.1.3.4 Perbandingan plot hasil cluster berdasarkan jumlah tifoid, umur, dan kondisi lingkungan sebelum dan sesudah optimasi.....	66
Gambar 4.1.3.5 Daftar desa pada cluster 1 berdasarkan jumlahTifoid dan umur.	71
Gambar 4.1.4. 1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah tb, umur, dan kebiasaan pola hidup	73
Gambar 4.1.4. 2 Perbandingan nilai centroid cluster berdasarkan jumlah TB, umur, dan kebiasaan pola hidup sebelum dan sesudah optimasi	75
Gambar 4.1.4.3 Sampel hasil cluster jumlah TB, umur, dan kebiasaan pola hidup setelah optimasi.....	75
Gambar 4.1.4.4 Perbandingan plot hasil cluster jumlah TB, umur, dan kebiasaan pola hidup.....	76
Gambar 4.1.4.5 Daftar desa pada cluster 1 berdasarkan jumlah TB, umur, dan kebiasaan pola hidup	77
Gambar 4.1.5 1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup.....	79
Gambar 4.1.5 2 Perbandingan nilai centroid cluster berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup sebelum dan sesudah optimasi	81
Gambar 4.1.5.3 Sampel hasil cluster berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup setelah optimasi	81
Gambar 4.1.5.4 Plot hasil cluster berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup.....	82
Gambar 4.1.5.5 Daftar desa pada cluster 1 berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup	83
Gambar 4.1.6.1 Grafik hubungan k terhadap SSE dan Silhouette Coefficient berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup.....	85
Gambar 4.1.6. 2 Perbandingan nilai centroid cluster berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup sebelum dan sesudah optimasi	86
Gambar 4.1.6.3 Sampel hasil cluster berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup setelah optimasi	87
Gambar 4.1.6.4 Perbandingan plot hasil cluster berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup sebelum dan sesudah optimasi.....	88
Gambar 4.1.6.5 Daftar desa pada cluster 1 berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup	89
Gambar 4.2.10.1 Tampilan halaman beranda WebGIS	137
Gambar 4.2.10. 2Tampilan halaman data	138
Gambar 4.2.10. 3 Tampilan halaman clustering	138
Gambar 4.2.10.4 Tampilan peta di halaman clustering	139
Gambar 4.2.10.5 Tampilan ketika salah satu cluster diklik dalam peta di halaman clustering.....	139

Gambar 4.2.10.6 Tampilan daftar cluster di halaman clustering	140
Gambar 4.2.10.7 Tampilan peta detail hasil cluster	140
Gambar 4.2.10.8 Tampilan daftar cluster di halaman clustering	141
Gambar 4.2.10.9 Tampilan peta ketika salah satu cluster di daftar cluster diklik	141
Gambar 4.2.10.10 Tampilan button di atas peta halaman detail clustering	142
Gambar 4.2.10.11 Tampilan interpretasi cluster di halaman clustering	142

DAFTAR TABEL

Tabel 4.1.1.1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah TB, umur, dan kondisi lingkungan	32
Tabel 4.1.1. 2 Nilai statistik cluster berdasarkan jumlah TB, umur, dan kondisi lingkungan.....	44
Tabel 4.1.2. 1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah DBD, umur, dan kondisi lingkungan.....	50
Tabel 4.1.2.2 Nilai statistik cluster berdasarkan jumlah DBD, umur, dan kondisi lingkungan.....	55
Tabel 4.1.3.1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah tifoid, umur, dan kondisi lingkungan.....	61
Tabel 4.1.3.2 Nilai statistik cluster berdasarkan jumlah tifoid, umur, dan kondisi lingkungan.....	66
Tabel 4.1.4.1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah tb, umur, dan kebiasaan pola hidup	73
Tabel 4.1.4.2 Nilai statistik cluster berdasarkan jumlah TB, umur, dan kebiasaan pola hidup.....	77
Tabel 4.1.5.1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup.....	79
Tabel 4.1.5.2 Nilai statistik cluster berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup.....	83
Tabel 4.1.6.1 Nilai k, SSE, Silhouette Coefficient berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup.....	85
Tabel 4.1.6.2 Nilai statistik cluster berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup	89
Tabel 4.2.1.1 Nilai SSE, Silhouette coefficient, dan DBI hasil cluster.....	91
Tabel 4.2.2.1 Hasil clustering berdasarkan jumlah TB, umur, dan kondisi lingkungan.....	92
Tabel 4.2.3.1 Hasil clustering berdasarkan jumlah DBD, umur, dan kondisi lingkungan.....	94
Tabel 4.2.4.1 Hasil clustering berdasarkan jumlah tifoid, umur, dan kondisi lingkungan.....	96
Tabel 4.2.5.1 Hasil clustering berdasarkan jumlah TB, umur, dan kebiasaan pola hidup.....	98
Tabel 4.2.6.1 Hasil clustering berdasarkan jumlah DBD, umur, dan kebiasaan pola hidup.....	100
Tabel 4.2.7.1 Hasil clustering berdasarkan jumlah tifoid, umur, dan kebiasaan pola hidup.....	101
Tabel 4.2.8.1 Interpretasi hasil clustering untuk setiap desa di Kabupaten Gowa	105

DAFTAR LAMPIRAN

Lampiran 1. Source Code.....	150
-------------------------------------	------------

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya, sehingga dapat menyelesaikan tugas akhir dengan judul “Optimasi K-means Clustering Pada Penyakit Menular Dengan Algoritma Particle Swarm Optimization” sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 di Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Penulis menyadari bahwa penyusunan dan penulisan tugas akhir ini tidak dapat terselesaikan dengan baik tanpa adanya bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih banyak kepada:

1. Allah SWT atas berkat dan Rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Kedua orang tua penulis, Bapak A.Rusli, S.Pd., M.M. dan Ibu Daryani Latif yang selalu memberikan dukungan, doa, dan semangat yang tiada hentinya, serta selalu sabar dalam mendidik penulis sejak kecil.
3. Ibu Elly Warni, S.T., M.T. selaku pembimbing I dan Bapak Ir. Christoforus Yohannes, M.T. selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
4. Segenap staf dan dosen Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir
5. Dr. Ridha, Ibu Hj. Eni, Ibu Hj. Relawati, Pak Ilham, Kak Fikri, dan seluruh staf Puskesmas Bontomarannu yang senantiasa memberikan nasihat dan masukan serta membantu penulis dalam pengumpulan data dalam penelitian ini.
6. Dea Wahsa Saputri, Putri Ayu Nengsih, Reskita Amelia, Mirnawati Darwis, A Diatri Nur Azkiyah ciwi-ciwi yang selalu menjadi partner dalam membantu dan mendengar curhatan penulis selama penulis menjadi mahasiswa sampai sekarang.
7. Nur Annisa Yusrah, Pahrul, Farhan, Ilal, Arif, Leon, Rayyan, dan teman-teman S19NIFIER yang selalu mendengar keluh kesah dan selalu membantu penulis saat penulis membutuhkan bantuan.
8. Teman-teman KKNT Posko Desa Toddopuli Gel.108 (Ryan, Fikar, Palli, Rahmat, Riana, Joya, Ririn, Dinda, Jeridh) yang telah memberikan pengalaman kepada penulis selama KKN.
9. Serta pihak-pihak lain yang tidak disebutkan dan tanpa sadar telah menjadi inspirasi dan membantu penulis dalam menyelesaikan tugas akhir

Penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan serta jasa dari semua pihak yang telah banyak membantu penulis dalam menyelesaikan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan segala bentuk saran serta

masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberi manfaat bagi para pembaca dan semua pihak. Aamiin.

Gowa, Mei 2023

Penulis,
Andi Rusmiati

BAB I PENDAHULUAN

1.1 Latar Belakang

Penyakit menular adalah penyakit yang dapat ditularkan melalui berbagai media. Penyakit jenis ini merupakan masalah kesehatan yang besar di hampir semua negara berkembang karena angka kesakitan dan kematiannya yang relatif tinggi dalam kurun waktu yang relatif singkat. Penyakit menular umumnya bersifat akut (mendadak) dan menyerang semua lapisan masyarakat. Penyakit jenis ini diprioritaskan mengingat sifat menularnya yang bisa menyebabkan wabah dan menimbulkan kerugian yang besar. Penyakit menular merupakan hasil perpaduan berbagai faktor yang saling mempengaruhi. Adapun yang termasuk penyakit menular, yakni Tuberkulosis (TBC), Demam Berdarah Dengue (DBD), dan Demam Tifoid (Widoyono, 2011).

Tuberkulosis (TB) adalah suatu penyakit menular yang disebabkan oleh kuman dari kelompok *Mycobacterium* yaitu *Mycobacterium tuberculosis*. Sumber penularan adalah pasien TB BTA positif melalui percik relik dahak yang dikeluarkannya (Kemenkes RI, 2014). Tuberkulosis (TB) adalah infeksi bakteri yang dapat menyerang hampir semua bagian tubuh, tetapi paling sering menyerang paru-paru, kondisi ini disebut ‘tuberkulosis paru-paru’ (Queensland Health, 2017).

Demam Tifoid atau *typhoid fever* adalah penyakit demam akut yang menular yang disebabkan akibat infeksi *Salmonella Typhi*, biasanya melalui konsumsi makanan atau air yang terkontaminasi. Penyakit akut ini ditandai oleh demam berkepanjangan, sakit kepala, mual, kehilangan nafsu makan, dan sembelit atau kadang-kadang diare. Dari data WHO didapatkan perkiraan jumlah kasus demam tifoid mencapai angka antara 11 dan 21 juta kasus dan 128.000 hingga 161.000 kematian terkait demam tifoid terjadi setiap tahun di seluruh dunia. Penyakit serupa tetapi seringkali kurang parah, demam paratifoid, disebabkan oleh *Salmonella Paratyphi* (WHO, 2018).

Demam berdarah dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue melalui gigitan nyamuk aedes terutama aedes aegypti. Menurut *World Health Organization* (WHO) hingga tahun 2007 DBD telah terjadi di 65 negara dengan laporan rata-rata kasus 925.896 per tahun. Negara beriklim tropis dan

subtropis berisiko tinggi terhadap penularan virus tersebut. Hal ini dikaitkan dengan kenaikan temperatur yang tinggi dan perubahan musim hujan dan kemarau disinyalir menjadi faktor risiko penularan virus *dengue* (Kemenkes, 2011).

Berdasarkan hasil riset yang telah dilakukan di beberapa puskesmas yang ada di kabupaten gowa, diperoleh bahwa pada tahun 2020 hingga 2022, Puskesmas Bontonompo II menangani sekitar 50 kasus tbc, 20 kasus DBD, dan 258 kasus demam tifoid. Puskesmas gowa menangani 212 kasus tbc, 140 kasus dbd, dan 377 kasus demam tifoid. Sedangkan Puskesmas Bontomarannu menangani 212 kasus tbc, 224 kasus DBD, dan 427 kasus demam tifoid. Dari data tersebut, dapat dilihat bahwa puskesmas bontomarannu menjadi salah satu puskesmas yang ada di Kabupaten Gowa yang menangani banyak kasus penyakit menular tbc, dbd, dan demam tifoid. Maka dari itu, pada penelitian ini, akan dilakukan proses clustering pada penyakit menular tersebut guna memperoleh informasi untuk dijadikan sebagai acuan atau bahan kebijakan bagi pihak puskesmas atau pemerintah guna melakukan pencegahan dari penyakit menular ini. Angka tersebut tentu bukan berita yang baik terutama bagi Puskesmas Bontomarannu yang merupakan salah satu puskesmas yang menangani banyak kasus penyakit menular sebab angka tersebut terus menerus bertambah akibat penularan yang terjadi. Tingginya kasus penyakit yang terjadi ini menyulitkan pihak puskesmas dalam mengelompokkan faktor-faktor apa yang menjadi penyebab bertambahnya penyakit tersebut setiap harinya karena salah satu alasannya adalah keterbatasan tenaga medis untuk mendeteksi secara manual indikator apa yang bisa dijadikan acuan melalui informasi pasien yang terkena penyakit tersebut.

Di Indonesia, banyak masyarakat yang masih menganggap penyakit menular ini sebagai penyakit yang kurang serius untuk ditangani sedangkan angka kematian yang diakibatkan oleh penyakit-penyakit ini terus bertambah setiap tahunnya. Tentu pemerintah dalam hal ini sudah melakukan usaha-usaha untuk meminimalisir masalah ini, salah satunya dengan memaksimalkan layanan-layanan kesehatan dan tenaga kesehatan, dan tentu juga aspek-aspek lain yang secara langsung/ tidak langsung sebagai faktor pendukung. Aspek lain yang bisa menjadi pendukung adalah terkait pengelolaan data, dimana kita bisa melihat terkadang data hanya digunakan sebagai bahan untuk menampilkan kondisi yang telah/sedang terjadi.

Namun, sebenarnya ada potensi yang bisa didapat dari data, yaitu bagaimana data tersebut bisa diolah untuk mendukung sebuah penyelesaian kasus. Salah satu contoh pemanfaatan dari data yaitu menganalisa data data tersebut sehingga bisa menghasilkan informasi yang berguna.

Dalam kasus penyakit menular ini, maka penulis akan melakukan pengelolaan data dari daftar pasien yang terkena penyakit tuberkulosis, demam tifoid, dan demam berdarah dengue di Puskesmas Bontomarannu. Salah satu teknik *clustering* yang bisa digunakan untuk mengelompokkan data yaitu *K-means clustering* yang akan mengelompokkan data yang memiliki karakteristik yang sama ke dalam satu *cluster* yang sama.

Penelitian yang dilakukan oleh Rozzi Kesuma Dinata (2020) menganalisis *K-means clustering* pada data sepeda motor mendapatkan nilai rata-rata *precision* sebesar 76%, *recall* sebesar 76%, dan *accuracy* sebesar 77% dari total 15 pengujian. Ari Yunus Hendrawan (2020) juga melakukan penelitian dengan meningkatkan kinerja algoritma *K-means* dengan menggunakan *Particle Swarm Optimization* dalam pengelompokan data penyediaan akses sanitasi dan air bersih dimana nilai *Bouldin Index* yang diperoleh dengan hanya menggunakan *K-means* yaitu sebesar 0.208856, sedangkan nilai *Bouldin Index* yang diperoleh dengan menerapkan PSO pada *K-means* sebesar 0.08383. Penelitian yang dilakukan oleh Khairati (2019) terkait Kajian Indeks Validitas pada Algoritma *K-Means Enhanced* dan *K-Means MMCA* mendapatkan hasil bahwa diantara indeks validitas *clustering* yakni *Silhouette*, *DBI*, *Calinski-Harabasz*, dan *Dunn*, disimpulkan bahwa indeks *Silhouette* dan *DBI* yang dapat memprediksi jumlah *cluster* optimal yang lebih baik dibanding indeks validitas yang lain.

Maka dari itu, penelitian ini akan mengembangkan penelitian yang telah dilakukan sebelumnya dengan melakukan pengoptimalan dari *K-means clustering* dengan mengoptimalkan nilai dari jarak *cluster* menggunakan algoritma *Particle Swarm Optimization* (PSO) dengan hasil akhir *clustering* akan divisualisasikan kedalam WebGIS (peta digital) untuk mempermudah pihak puskesmas dalam membaca data hasil *clustering*. Hasil optimasi *cluster* diperoleh dengan melihat nilai indeks validitas *cluster* yakni, *Sum of Square Error* (SSE), *Silhouette Coefficient*, dan *Davies Bouldin Index* (DBI). Pemanfaatan dari *clustering* ini yaitu

dapat menjadi solusi dalam membantu analisis penyakit menular yakni tuberkulosis, DBD, dan demam tifoid pada masyarakat berdasarkan variabel-variabel penelitian. Diharapkan dengan adanya penelitian ini dapat memberikan informasi sebagai dasar untuk melakukan tindakan yang diperlukan untuk pengobatan, pengendalian, dan pencegahan.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana melakukan optimasi *K-means clustering* dengan menggunakan algoritma PSO?
2. Bagaimana hasil *clustering* data pasien tuberkulosis, DBD, dan demam tifoid per desa di Puskesmas Bontomarannu?
3. Bagaimana memvisualisasikan hasil *clustering* ke dalam WebGIS?

1.3. Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah sebagai berikut:

1. Untuk mendapatkan hasil yang lebih optimal dari *K-means clustering*
2. Untuk membuat *clustering* data pasien tuberkulosis, DBD, dan demam tifoid per desa di Puskesmas Bontomarannu
3. Untuk membuat sistem WebGIS yang akan memvisualisasikan hasil *clustering* ke dalam bentuk peta digital

1.4. Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah dapat memberikan informasi sebagai dasar dalam pengambilan keputusan dan tindakan untuk pengendalian dan pencegahan kasus penyakit tuberkulosis, demam tifoid, dan demam berdarah dengue di masyarakat.

1.5. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Pengambilan data dilakukan di Puskesmas Bontomarannu, yang berupa kuesioner yang dibagikan ke pasien tuberkulosis, DBD, dan demam tifoid yang kemudian akan dibentuk dalam format *excel* (.csv).
2. Metode *clustering* yang digunakan adalah *K-means Clustering* dengan mengoptimalkan titik *centroid* menggunakan algoritma PSO.

1.6. Sistematika Penulisan

Berikut adalah gambaran singkat mengenai isi tulisan secara keseluruhan:

BAB I PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas landasan teori yang digunakan untuk menganalisis masalah yang akan diteliti serta hal-hal lain yang berhubungan dengan variabel-variabel data yang akan digunakan, seperti tuberkulosis, demam tifoid, demam berdarah dengue, data mining, *clustering*, algoritma *K-means*, serta WebGIS.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi mengenai tahapan penelitian, instrumen penelitian, pengumpulan data, penerapan algoritma, teknik pengolahan data, serta visualisasi hasil *clustering* pada WebGIS.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang sistem yang telah dibangun serta pembahasan dari hasil penelitian.

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.

BAB II

TINJAUAN PUSTAKA

2.1 Penyakit Menular

Menurut Peraturan Menteri Kesehatan Nomor 82 Tahun 2014, Penyakit menular adalah penyakit yang menular ke semua umur yang disebabkan oleh agen biologi, antara lain virus, bakteri, jamur, dan parasit. Penyakit menular dapat ditularkan dari satu orang ke orang yang lain, dari orang ke hewan, ataupun sebaliknya. Keberadaan virus, bakteri, jamur, dan/atau parasit di dalam atau di permukaan tubuh dapat menyebabkan infeksi. Perpindahan agen infeksi atau parasit tersebut dari individu yang sakit ke individu yang sehat dinamakan sebagai penularan penyakit. Penularan penyakit ini bisa terjadi secara langsung atau tidak langsung dan juga bisa menular melalui perantara atau penghubung. Penyakit menular biasanya ditandai dengan adanya penyebab penyakit yang hidup dan bisa berpindah menyerang inang (penderita).

Bagi Indonesia sebagai negara berkembang, permasalahan penyakit menular masih menjadi permasalahan rutin setiap tahunnya karena seringkali menimbulkan Kejadian Luar Biasa (KLB) sehingga menjadi beban bagi pemerintah dan kesengsaraan bagi masyarakat. Adapun penyakit menular di Indonesia diklasifikasikan dalam tiga kelompok yakni penyakit menular langsung, penyakit yang ditularkan melalui vektor dan zoonosis, serta penyakit yang dapat dikendalikan dengan imunisasi. Penularan penyakit secara langsung terjadi ketika kuman pada orang yang sakit berpindah melalui kontak fisik, misalnya lewat sentuhan, melalui udara saat bersin dan batuk, atau melalui kontak dengan cairan tubuh seperti *urine* dan darah. Adapun orang yang menularkan penyakit bisa saja tidak memperlihatkan gejala dan tidak terlihat seperti orang sakit, apabila mereka hanya sebagai pembawa (*carrier*) penyakit (Hulu *et al.*, 2020).

2.2 Tuberkulosis

Tuberkulosis (TB) adalah penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis* dan disebut sebagai Bakteri Tahan Asam (BTA) (Infodatin Kemenkes RI, 2018). Sebagian besar bakteri TB menyerang paru

(TB paru), namun dapat juga mengenai organ tubuh lainnya (TB ekstra paru). Penularan TB terutama terjadi secara aerogen atau lewat udara dalam bentuk *droplet* (percikan dahak/sputum). Sumber penularan TB yaitu penderita TB BTA positif ketika batuk, bersin, atau berbicara mengeluarkan *droplet* yang mengandung bakteri *M. tuberculosis* (Kemenkes RI, 2017).

Menurut *World Health Organization* (2022), Kasus tuberculosis tertinggi berada di lima negara, yaitu India, China, Philipina, Pakistan, dan Indonesia. Indonesia menjadi negara tertinggi ketiga setelah India dan China dengan jumlah kasus tertinggi tuberculosis di dunia. Jumlah kasus tuberculosis di Indonesia pada tahun 2022 meningkat dari tahun sebelumnya sebesar 717.941 kasus *Case Notification Rate* (CNR) 354 per 100.000 penduduk, dimana tahun 2021 sebesar 443.235 kasus dengan CNR 214 per 100.000 penduduk. Hasil ini menunjukkan jumlah kasus *tuberculosis* melonjak naik sebanyak 61.98%.

Tuberculosis sebagian besar menyerang orang dewasa di tahun-tahun paling produktif. Namun, semua kelompok umur berisiko. Lebih dari 95% kasus dan kematian terjadi di negara berkembang. Orang yang terinfeksi HIV 18 kali lebih mungkin mengembangkan TB aktif. Risiko TB aktif juga lebih besar pada orang yang menderita kondisi lain yang merusak sistem kekebalan. Orang dengan kekurangan gizi 3 kali lebih berisiko. Secara global pada tahun 2019 terdapat 2,2 juta kasus baru TB pada tahun 2018 yang disebabkan oleh gizi kurang (WHO, 2020).

Penyebab tuberculosis dipengaruhi oleh berbagai faktor. Faktor pertama yang mempengaruhi adalah Umur, di mana terdapat insiden tertinggi penyakit tuberculosis pada Umur dewasa muda, dan diperkirakan sekitar 75% dari penderita tuberculosis berada pada kelompok Umur produktif di Indonesia. Faktor kedua adalah jenis kelamin, di mana lebih banyak laki-laki yang terkena penyakit ini karena kebanyakan dari mereka memiliki kebiasaan merokok. Faktor ketiga adalah kebiasaan merokok, yang dapat melemahkan sistem kekebalan tubuh sehingga meningkatkan risiko terkena penyakit, terutama pada laki-laki yang merokok. Faktor keempat adalah kepadatan hunian, yang merupakan faktor lingkungan penting, terutama pada penderita tuberculosis,

karena kuman *M. Tuberculosis* dapat tumbuh dan menyebar dengan mudah di lingkungan yang gelap dan tidak terkena sinar matahari. Faktor kelima adalah pekerjaan, di mana orang yang bekerja di sektor kesehatan berisiko tinggi terkena penyakit ini melalui kontak langsung dengan pasien, namun masih ada beberapa pekerjaan lain yang dapat meningkatkan risiko. Faktor keenam adalah status ekonomi, yang merupakan faktor utama dalam keluarga dengan pendapatan rendah. Orang dengan pendapatan rendah cenderung lebih rentan terkena penyakit karena tidak mampu memenuhi kebutuhan kesehatan yang memadai (Luthfi, 2012).

2.3 Demam Tifoid

Demam tifoid adalah suatu penyakit sistemik yang bersifat akut yang disebabkan oleh kuman berbentuk basil yaitu *Salmonella typhi* yang ditularkan melalui makanan dan minuman yang tercemar feses. Demam tifoid merupakan salah satu dari masalah kesehatan masyarakat yang penting karena penyebab penyakit tifoid berkaitan dengan kepadatan penduduk, kesehatan lingkungan, sumber air yang buruk, serta pengelolaan makanan yang masih rendah dari kebersihan. Demam tifoid mudah dikenali dengan gejala awal seperti demam dengan suhu tinggi, nyeri perut, sakit kepala, dan kadang sembelit pada orang dewasa dan diare pada anak-anak (Kemenkes, 2006).

Demam tifoid terdapat di seluruh dunia, terutama pada negara-negara yang berkembang dan di daerah tropis. Demam tifoid di Indonesia jarang dijumpai secara epidemi tetapi bersifat endemis dan banyak dijumpai di kota-kota besar. Insiden penyakit tifoid masih sangat tinggi kisaran 350 - 810 per 100.000 penduduk, kasus di rumah sakit cenderung meningkat setiap tahunnya. 500 per 100.000 penduduk serta angka kematian diperkirakan sekitar 0,6% - 5% akibat dari keterlambatan mendapatkan pengobatan (Kemenkes, 2006).

Beberapa gejala demam tifoid yang sering terjadi adalah sebagai berikut.

a. Demam

Demam adalah gejala utama dari tifoid. Penyakit ini diawali dengan suhu tubuh yang sering naik turun, pada pagi hari lebih rendah atau normal, sore dan malam makin tinggi dan setiap hari demam semakin tinggi dan

disertai banyak gejala seperti kepala pusing, nyeri otot, pegal-pegal, insomnia, mual, dan muntah (Kemenkes, 2006).

b. Gangguan Saluran Pencernaan

Gejala lainnya dari demam tifoid ini juga ditandai dengan adanya bau mulut yang tidak enak karena demam yang lama, bibir kering dan pecah-pecah, serta lidah terlihat kotor dan ditutupi selaput putih, umumnya penderita sering mengeluh nyeri perut terutama pada ulu hati disertai mual dan muntah, awal sakit sering konstipasi dan pada minggu berikutnya muncul diare (Kemenkes, 2006).

c. Gangguan Kesadaran

Gejala yang ketiga yaitu umumnya terdapat gangguan kesadaran yang ringan dan sering didapat kesadaran apatis dengan kesadaran seperti berkabut. Bila sudah berat, terkadang penderita bisa sampai koma dengan toksik (Kemenkes, 2006).

d. Hepatosplenomegali

Gejala lainnya yang sering terjadi bagi seseorang yang terkena demam tifoid yaitu hepatosplenomegali dimana hati dan atau limpa mengalami pembesaran. Hati juga terasa kenyal dan nyeri saat ditekan. Gejala ini terjadi dikarenakan kuman penyebab infeksi masuk ke dalam hepar dimana akan mengeluarkan endotoksin yang akan merusak hepar sehingga menyebabkan terjadinya hepatomegali serta mengakibatkan splenomegali yang disertai dengan peningkatan SGOT atau SGPT (Kemenkes, 2006).

2.4 Demam Berdarah Dengue

Demam Berdarah Dengue (DBD) adalah penyakit menular yang disebabkan oleh virus yang ditularkan melalui gigitan nyamuk *Aedes aegypti*. DBD merupakan penyakit febril akut yang disertai dengan gejala pusing, nyeri otot dan tulang atau sendi, bercak-bercak merah dikulit dan leukopenia. DBD dapat dikenal dengan munculnya 4 gejala spesifik yaitu demam tinggi, hemoragic. hepatomegaly, dan kasus yang paling parah ditandai dengan adanya kegagalan pernafasan (WHO, 1997).

Kasus DBD banyak ditemukan di daerah tropis dan sub tropis. Data dari WHO menunjukkan Asia menjadi negara dengan urutan pertama kasus tertinggi DBD setiap tahunnya. Salah satu negara di Asia Tenggara yang menduduki peringkat pertama kasus DBD dari tahun 1969 sampai tahun 2009 adalah negara Indonesia. Di Indonesia, penyakit DBD mulai dikenal pertama kali tahun 1969 di Surabaya dengan jumlah kasus 58 orang, 24 orang diantaranya meninggal dunia (Angka Kematian (AK): 41,3%). Berdasarkan laporan Kementerian Kesehatan (Kemenkes), jumlah kasus DBD kumulatif tercatat sebanyak 131.265 kasus DBD pada tahun 2022 dimana sekitar 40% adalah anak-anak Umur 0-14 tahun. Sementara, jumlah kematian akibat DBD sebanyak 1.135 kasus dengan 73% terjadi pada anak Umur 0-14 tahun.

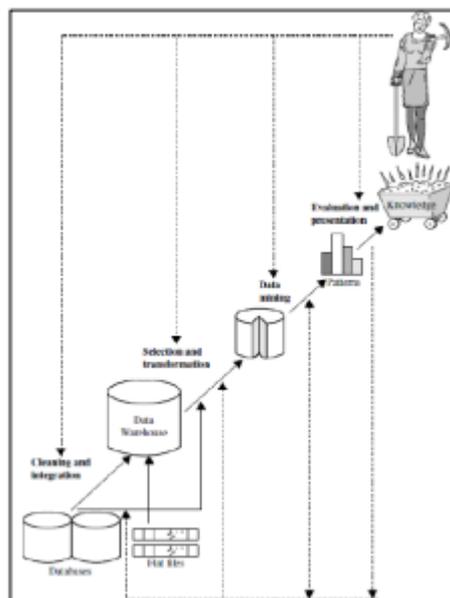
Epidemiologi DBD dapat berhubungan dengan wilayah, umur, serta jenis kelamin dimana pada tahun 2020 presentasi penderita laki-laki dan perempuan hampir sama yakni jumlah penderita berjenis kelamin laki-laki sebanyak 53,11% dan perempuan sebanyak 46,23%. Dapat disimpulkan bahwa resiko demam berdarah pada laki-laki dan perempuan hampir sama dan tidak tergantung pada jenis kelamin. Kemudian kasus demam berdarah berdasarkan umur pada tahun 2022 kelompok umur terbesar kasus demam berdarah cenderung pada kelompok umur di antara 0 - 14 tahun (Kemenkes, 2022).

Penularan DBD dapat terjadi disemua tempat yang terdapat vektor penularnya dalam hal ini vektor utamanya adalah nyamuk *Aedes aegypti*, kemudian tempat yang berpotensi untuk terjadinya penularan DBD seperti daerah yang menjadi wabah penyakit DBD, tempat umum yang merupakan tempat berkumpulnya orang dari berbagai tempat dan wilayah sehingga kemungkinan terjadinya pertukaran beberapa tipe virus dengue besar seperti sekolah, pasar, hotel, puskesmas, rumah sakit dan sebagainya dan pemukiman baru, karena di lokasi ini penduduk umumnya berasal dari berbagai wilayah, maka ada kemungkinan di antara pendatang baru tersebut terdapat karier tipe virus yang berlainan dari masing masing tempat asalnya (Widoyono, 2011).

2.5 Data Mining

Data mining adalah sebuah proses pencarian secara otomatis untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data (Ordila *et al.*, 2020).

Turban *et al.* (2018) menyebutkan bahwa *Knowledge Discovery in Database Process* (KDD) adalah salah satu metode yang bisa digunakan dalam melakukan data mining. Fayyad *et al.* (1996) mendefinisikan KDD sebagai proses dari menggunakan metode data mining untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data. Tujuan dari KDD yaitu untuk mencari potensi dari data-data yang diperoleh dari *database* yang nantinya akan dipelajari melalui pola yang kemudian akan dianalisis serta divisualisasikan sehingga mudah dipahami oleh pengguna. Proses KDD ini terdiri dari beberapa langkah, mulai dari *preprocessing* data hingga *postprocessing* hasil dari data mining.



Gambar 2. 1 Tahap data mining penemuan KDD (Han et al., 2011)

Jiawei Han et al (2011) menyebutkan bahwa data mining adalah bagian internal dari *knowledge discovery in databases* (KDD) yakni sebuah langkah dalam proses mencari sebuah pola yang terdapat dalam setiap informasi. Adapun langkah-langkah tersebut yakni sebagai berikut.

a. *Data cleaning*

Tahap untuk menghapus data yang tidak dipakai dan data yang tidak konsisten.

b. *Data integration*

Tahap untuk menggabungkan beberapa sumber data.

c. *Data selection*

Tahap dimana data yang bersangkutan pada tugas analisis diseleksi dan diambil kembali dari *database*.

d. *Data transformation*

Tahap untuk mengubah atau memperkuat data menjadi bentuk yang seharusnya untuk diolah dengan menganalisis ringkasan atau jumlah total agregasi.

e. *Data mining*

Tahap penting dimana metode intelijen diterapkan dengan tujuan untuk mengolah pola-pola data.

f. *Pattern evaluation*

Tahap untuk mengidentifikasi pola-pola yang menarik yang menjelaskan mengenai ukuran dasar pengetahuan yang ada..

g. *Knowledge presentation*

Tahap visualisasi dan teknik representasi *knowledge* yang digunakan untuk menyajikan *knowledge* yang telah diolah untuk pengguna.

2.6 K-means Clustering

K-means clustering merupakan algoritma pengelompokan *iterative* yang melakukan partisi set data ke dalam sejumlah *K cluster* yang telah ditetapkan di awal. Wu dan Kumar (2009) menyebutkan bahwa secara historis, *K-means clustering* menjadi salah satu algoritma yang paling penting dalam bidang *data mining*. Secara sederhana, *K-means* diimplementasikan dan dijalankan, umum

penggunaannya dalam praktek, mudah beradaptasi, dan *relative* cepat. Algoritma *K-means* termasuk dalam metode *non-hierarchical* yang akan mempartisi data ke dalam bentuk satu atau lebih *cluster*, sehingga data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam satu *cluster* yang sama, serta data yang memiliki karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* yang lainnya.

Jiawei Han *et al* (2012) menyebutkan bahwa algoritma *K-means* merupakan algoritma yang menerapkan teknik pengelompokan berbasis titik pusat (*Centroid-Based Technique*). Adapun langkah-langkah dari algoritma *K-means* adalah sebagai berikut (Parteek Bhatia, 2019).

- a. Menentukan jumlah *cluster* awal yang ingin dibentuk (k).
- b. Mengidentifikasi nilai *centroid*.

Penentuan nilai awal dari *centroid* diambil dari data yang ada yang diambil secara acak atau *random* untuk awal iterasi. Ketika menghitung nilai dari *centroid* yang merupakan tahap dari iterasi, digunakan rumus sebagai berikut.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$$

dimana,

v_{ij} : *centroid* atau *mean cluster* ke- i untuk variabel ke- j

N_i : jumlah data yang menjadi anggota *cluster* ke- i

i, k : indeks dari *cluster*

j : indeks dari variabel

x_{kj} : nilai data ke- k yang ada di dalam *cluster* tersebut untuk variabel ke- j

- c. Menghitung jarak antar *centroid* dengan titik tiap objek agar ditemukan jarak terdekat dari setiap data dengan menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

dimana,

D_e : *Euclidean Distance*

i : banyak objek

(x, y) : koordinat objek

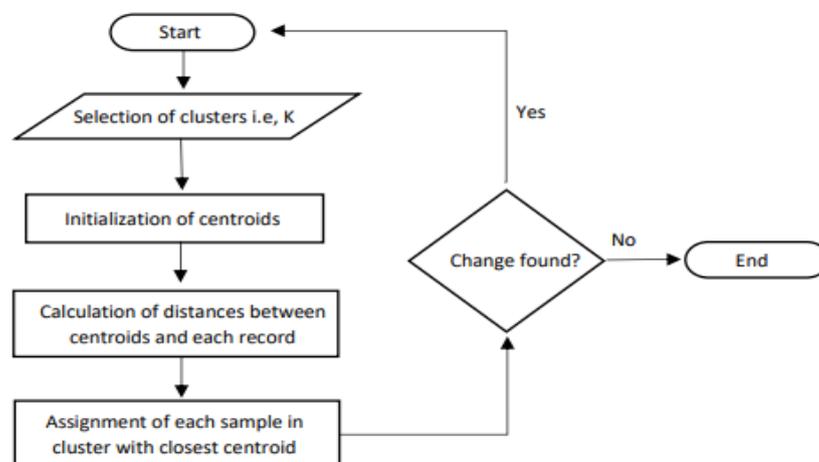
(s, t) : koordinat *centroid*

d. Pengelompokan objek

Menghitung jarak minimum objek untuk menentukan anggota dari *cluster*. Nilai yang diperoleh dalam keanggotaan data pada *distance* matriks adalah 0 atau 1, dimana nilai 0 untuk data yang dialokasikan ke *cluster* lain dan nilai 1 untuk data yang dialokasikan ke *cluster*. Pengelompokan objek berdasarkan dengan jarak terdekat dengan *centroid*.

e. Penentuan pusat *cluster* baru

Menentukan nilai *centroid* pada tahap iterasi agar diperoleh nilai *centroid* yang baru. Lakukan iterasi hingga nilai *centroid* yang diperoleh tidak berubah dan anggota *cluster* tidak berpindah ke *cluster* yang lain sehingga proses selesai.



Gambar 2. 2 *Flowchart K-means clustering* (Parteek Bhatia, 2019)

K-means dapat diterapkan pada data yang direpresentasikan dalam r -dimensi ruang tempat dimana *K-means* akan mengelompokkan set data r -

dimensi, $X = \{x_i | i = 1, \dots, N\}$, dimana $x_i \in R^d$ yang menyatakan data ke- i sebagai “titik data”. *K-means* akan mempartisi X ke dalam K *cluster*. Algoritma *K-means* akan mengelompokkan semua titik data dalam X sehingga setiap titik x_i hanya jatuh ke dalam satu dari K partisi.

Kotu *et al* (2015) menyebutkan bahwa *clustering K-means* merupakan metode pengelompokan berbasis *prototype* dimana kumpulan data dibagi menjadi *cluster* k . Dalam teknik ini, pengguna awalnya menentukan jumlah *cluster* (k) yang perlu dikelompokkan ke dalam kumpulan data. Adapun tujuan dari *clustering K-means* yaitu untuk menemukan titik *prototype* data untuk setiap *cluster*. Semua titik data tersebut kemudian akan ditugaskan untuk *prototype* terdekat, yang kemudian akan membentuk *cluster*. Pusat dari *cluster* dapat menjadi *mean* dari semua objek dalam *cluster* seperti pada *K-means*, atau objek data yang mewakili, seperti pada *clustering k-medoid*.

2.7 Particle Swarm Optimization

Algoritma *Particle Swarm Optimization* (PSO) merupakan salah satu teknik komputasi *evolusioner* yang mempunyai kesamaan dengan *Genetic Algorithm*, yaitu algoritma ini dimulai dengan membangkitkan populasi secara acak atau *random*. *Particle Swarm Optimization* adalah suatu teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter (Basari *et al.*, 2013).

Liu *et al* (2011) menyebutkan bahwa PSO banyak digunakan untuk memecahkan masalah optimasi serta sebagai masalah seleksi fitur, Algoritma dalam algoritma PSO, terdapat beberapa cara untuk melakukan pengoptimasian, yaitu dengan meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, serta menyeleksi atribut (*attribute selection*) dan *feature selection*.

Untuk menemukan solusi yang optimal, masing-masing partikel bergerak ke arah posisi sebelumnya terbaik (*pbest*) dan terbaik posisi global (*gbest*). Kecepatan dan posisi partikel dapat dipengaruhi dengan persamaan:

$$V_{ij}(t + 1) = w * V_{ij}(t) + c_1 * rand_1 * (pbest_{ij}(t) - p_{ij}(t)) + c_2 * rand_2 * (gbest_{ij}(t) - p_{ij}(t))$$

$$p_{ij}^t = p_{ij}^{t-1} + V_{ij}^t$$

dimana,

t : menunjukkan *counter* iterasi

V_{ij} : kecepatan partikel i pada dimensi ke- j

p_{ij} : posisi partikel i pada dimensi ke- j

$Pbest_{ij}$: posisi pbest partikel i pada dimensi ke- j

$Gbest_{ij}$: posisi gbest dari dimensi ke- j

w : bobot inersia

$rand_1, rand_2$: fungsi acak rentang [0, 1]

c_1 dan c_2 : koefisien percepatan positif

2.8 Sum of Squared Error (SSE)

Sum of Squared Error (SSE) adalah perbedaan kuadrat antara setiap pengamatan dan rata-rata (*mean*) kelompoknya. SSE dapat digunakan sebagai ukuran variasi dalam sebuah *cluster*. Jika semua kasus dalam *cluster* identik, maka SSE akan sama dengan 0. SSE berfungsi dalam mengukur kualitas yang merupakan bagian dari *cluster* serta kualitas hasil dari penjumlahan seluruh jarak masing-masing data dengan titik pusat *cluster*-nya. Semakin kecil nilai dari SSE, maka semakin tinggi tingkat similaritas data yang ada di dalam masing-masing *cluster* atau semakin baik *cluster* yang diperoleh. Adapun persamaan dari SSE menurut Tippaya Thinsungnoen *et al* (2015) yaitu sebagai berikut.

$$SSE = \sum_{i=1}^n (d_i - c_i)^2$$

dimana,

SSE : nilai kuadrat selisih antara koordinat *centroid* setiap data

n : jumlah data

d_i : nilai data ke- i

c_i : nilai *centroid* ke- i

2.9 Silhouette Coefficient

Silhouette Coefficient adalah suatu metode untuk melakukan evaluasi atau menguji kualitas dari *cluster* yang diperoleh atau dihasilkan yang akan menggabungkan antara metode *cohesion* dan *separation*. Hasil akhir dari *Silhouette Coefficient* yaitu angka antara -1 hingga 1. Apabila nilai dari *Silhouette Coefficient* mendekati angka 1, maka pengelompokan pada suatu *cluster* semakin baik, namun jika mendekati angka -1, maka pengelompokan *cluster* tersebut semakin buruk.

Nilai *Silhouette Coefficient* untuk setiap data dapat didefinisikan dengan persamaan berikut.

$$s(i) = \frac{(b_i - a_i)}{\max(a_i, b_i)}$$

Dimana $s(i)$ adalah nilai *Silhouette Coefficient*, a_i adalah jarak antara titik data dengan titik pusat *cluster*-nya, b_i adalah jarak antara titik data dengan titik pusat *cluster* terdekat berikutnya. Untuk menilai suatu nilai *Silhouette Coefficient*, dapat dilihat pada table 2.1 yang telah dibuat oleh Kaufman dan Rousseeuw berikut. (Azuri *et al.*, 2016).

Tabel 2.1 Interpretasi Nilai *Silhouette Coefficient* (Azuri *et al.*, 2016)

Nilai <i>Silhouette Coefficient</i>	Interpretasi
0.71 – 1.00	Klaster yang kuat
0.51 – 0.70	Klaster telah baik atau sesuai
0.26 – 0.50	Klaster yang lemah
≤ 0.25	Tidak dapat dikatakan sebagai klaster

2.10 Davies-Bouldin Index

Metode *Davies-Bouldin Index* merupakan metode untuk mengevaluasi *cluster* yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin dan diberi nama *Davis-Bouldin Index* (DBI). Teknik ini dapat digunakan untuk mengukur kevalidan *cluster* pada metode klastering berbasis partisi, dengan menggunakan nilai kohesi dan separasi sebagai dasar. Nilai kohesi didefinisikan sebagai total kedekatan data dengan *centroid* dari *cluster* yang bersangkutan, sementara nilai separasi didasarkan pada jarak antara data dan *centroid* di luar *cluster* yang bersangkutan (Permatadevi *et al.*, 2013).

Dalam DBI, *Sum of Square Within Cluster* (SSW) adalah persamaan yang digunakan untuk menentukan nilai matriks kohesi pada sebuah *cluster* ke- i . Untuk menghitung nilai SSW, dapat digunakan rumus sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

Sementara *Sum of Square Between Cluster* (SSB) adalah persamaan yang digunakan untuk menentukan nilai separasi antara cluster. Nilai SSB dapat dihitung menggunakan rumus berikut:

$$SSB_{i,j} = d(c_i, c_j)$$

Setelah nilai kohesi dan separasi diperoleh, dilakukan pengukuran rasio (ij) untuk menentukan nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . *Cluster* yang dianggap baik adalah *cluster* yang memiliki nilai kohesi seminimal mungkin dan separasi semaksimal mungkin. Untuk menghitung nilai rasio, digunakan rumus sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Nilai rasio yang dihasilkan digunakan untuk mencari nilai *Davies-Bouldin Index* (DBI) menggunakan rumus berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

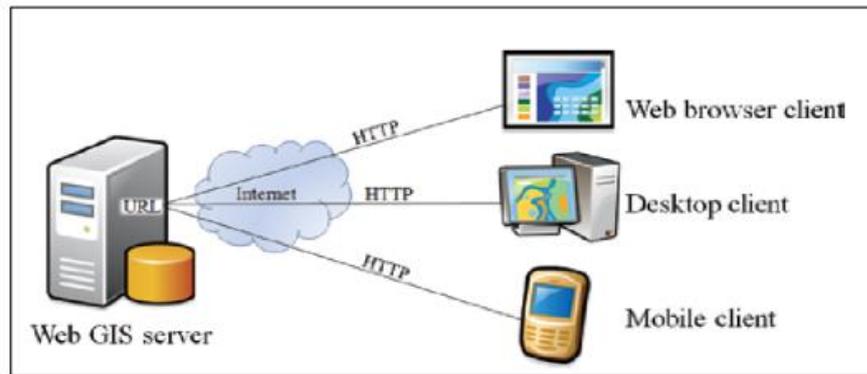
Hasil dari rumus tersebut menjadi ukuran validitas dari cluster yang sedang diuji. Semakin kecil nilai DBI yang dihasilkan (non-negatif dan ≥ 0), maka semakin baik *cluster* yang dihasilkan dari metode *clustering* yang digunakan.

2.11 Web GIS

Menurut Purwadhi (1994), Sistem Informasi Geografis atau *Geographics Information Sistem* adalah suatu sistem yang mengorganisir perangkat keras (*hardware*), perangkat lunak (*software*), dan data. Selain itu, sistem informasi geografis juga dapat mendayagunakan sebuah sistem penyimpanan, pengolahan, dan juga analisis data secara simultan agar dapat diperoleh informasi yang berkaitan dengan aspek keruangan. Selain itu, Chen *et al* (1999) mengatakan bahwa sistem informasi geografis adalah sebuah sistem komputer yang memungkinkan pengguna (*user*) untuk mengumpulkan, mengelola, dan menganalisis data spasial dalam ukuran yang besar.

WebGIS adalah aplikasi *Geografis Information Sistem* berbasis *Website* atau pemetaan digital yang memanfaatkan jaringan internet sebagai media komunikasi yang berfungsi mendistribusikan, mempublikasikan, mengintegrasikan, mengkomunikasikan, dan menyediakan informasi dalam bentuk teks, peta digital serta menjalankan fungsi-fungsi analisis dan *query* yang terkait dengan *GIS* melalui jaringan internet.

Menurut Fu (2011), *WebGIS* adalah setiap *GIS* yang menggunakan teknologi *web* untuk berkomunikasi antar sistem. *WebGIS* merupakan jenis sistem informasi yang terdistribusi. Bentuk paling sederhana dari *WebGIS* yaitu harus terdiri setidaknya dari *server* dan *client*. *Server* akan berperan sebagai aplikasi *server* berupa *URL* yang bisa diakses oleh *client* dengan menggunakan *HTTP*, sedangkan *client* dapat berupa *web browser*, aplikasi desktop, maupun aplikasi *celular*. Respon yang akan diberikan oleh *server* untuk ditampilkan kepada *client* dapat berupa *HTML*, gambar, *XML* (*Extensible Markup Language*), atau *JSON* (*JavaScript Object Notation*).



Gambar 2. 3 Arsitektur *WebGIS* (Fu, 2011)

WebGIS menyediakan mekanisme dan metode baru yang efektif dalam pengembangan sistem informasi geografis (SIG). Adapun arsitektur *WebGIS* terdiri dari tiga lapisan, yaitu lapisan *user interface*, lapisan *application server*, dan lapisan *database*. Zhelu (2009) menyatakan bahwa *Google Maps*, *Google Earth*, *OpenStreet Maps*, *Yahoo Maps*, *Microsoft's live search maps* dan banyak aplikasi komersial dan non-komersial lainnya menyediakan berbagai jenis informasi terkait geografis. Informasi yang dimaksud dapat berupa peta yang terperinci, citra satelis, dan peta daerah yang mencakup seluruh dunia dan memungkinkan pengguna untuk menggunakan *Application Programming Interface (API)*.