

**OPTIMASI K-MEANS CLUSTERING PADA PENGELOMPOKAN GEJALA  
DEMAM TIFOID DENGAN ALGORITMA PRINCIPAL COMPONENT  
ANALYSIS DAN PARTICLE SWARM OPTIMIZATION**



**PROPOSAL TUGAS AKHIR**

**Disusun Oleh:**

**ANDI RUSMIATI**

**D121191079**

**DEPARTEMEN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS HASANUDDIN**

**GOWA**

**2022**

## **LEMBAR PENGESAHAN**

Judul : “Optimasi K-Means Clustering dalam Pengelompokan Gejala Demam Tifoid dengan Algoritma Principal Component Analysis dan Particle Swarm Optimization”

Nama : Andi Rusmiati

NIM : D121191079

Departemen : Teknik Informatika

Diajukan sebagai salah satu syarat akademik pada Program Sarjana Universitas Hasanuddin.

**Menyetujui,**

**Pembimbing**

A handwritten signature in black ink, appearing to be 'Elly Warni'.

**Elly Warni, S.T., M.T.**

## **I. Judul**

Judul tugas akhir yang akan diangkat adalah “Optimasi K-Means Clustering dalam Pengelompokan Gejala Demam Tifoid dengan Algoritma Principal Component Analysis dan Particle Swarm Optimization”

## **II. Latar Belakang**

Demam Tifoid atau *typhoid fever* adalah penyakit demam akut 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).

Di Indonesia, banyak masyarakat yang masih menganggap demam tifoid hanya demam biasa dan tidak ingin melakukan pemeriksaan secara lebih intensif untuk lebih mengetahui lagi tentang gejala-gejala yang dirasakan. Kurangnya pengetahuan yang ada pada masyarakat mengenai penyakit demam tifoid merupakan salah satu alasan penyebab tingginya angka kematian akibat penyakit demam tifoid dan juga merupakan suatu permasalahan pada kasus demam tifoid ini. 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 demam tifoid ini, maka penulis akan melakukan pengelolaan data dari daftar pasien yang terkena demam tifoid disertai dengan gejala-gejala yang dialaminya, dimana data ini akan diolah untuk menghasilkan klasifikasi/pengelompokan gejala demam tifoid. 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 [1] 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. Dyah Hedyati [2] menerapkan PCA untuk reduksi dimensi pada proses clustering data produksi pertanian di Kabupaten Bojonegoro mendapatkan hasil yakni evaluasi data hasil clustering menggunakan nilai DB index menunjukkan nilai paling optimal pada dataset yang direduksi yaitu 0.4072. Sedangkan hasil clustering dataset tanpa proses reduksi dimensi menghasilkan nilai DB index 0.4598. Ari Yunus Hendrawan [3] 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.

Maka dari itu, penelitian ini akan mengembangkan penelitian yang telah dilakukan sebelumnya dengan melakukan pengoptimalan dari K-means clustering dengan melakukan reduksi dimensi menggunakan algoritma *Principal Component Analysis* (PCA) dan mengoptimalkan nilai dari jarak cluster menggunakan algoritma *Particle Swarm Optimization* (PSO). Pemanfaatan dari clustering ini yaitu dapat menjadi solusi dalam membantu analisis gejala 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.

### III. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah:

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

#### **IV. Tujuan**

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

#### **V. Manfaat**

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

#### **VI. Batasan Masalah**

Batasan masalah dalam penelitian ini adalah:

1. Data yang dianalisis dalam penelitian ini adalah data pasien yang terdeteksi demam tifoid di Puskesmas Bontomarannu tahun 2018 - 2022.
2. Variabel - variabel dalam penelitian terdiri dari:
  - Tanggal Kunjungan
  - Nama
  - Jenis Kelamin
    - Laki-Laki
    - Perempuan
  - Umur
  - Alamat
  - Gejala yang dirasakan:
    - Demam
    - Panas Dingin
    - Mual
    - Sakit Kepala
    - Lidah Kotor
    - Keringat Berlebih
    - Batuk
  - Status Tifoid:
    - Positif
    - Negatif

Tanggal Pengukuran	Nama	Jenis Kelamin			Usia (Tahun)	Alamat	Gejala								Status	
		Laki-Laki	Pemempuan				Demam	Panas Dingin	Mual	Sakit Kepala	Lidah Kotor	Keringat Berlelehan	Batuk Kering	Pilek	Ngalir	
3 Februari 2019	Karina	0	1	52	Sekeloa		1	0	1	1	0	0	0	1	0	
3 Februari 2019	Mulati	0	1	20	Sekeloa		1	1	1	1	0	0	0	1	0	
3 Februari 2019	Martani	0	1	31	Sekeloa		0	1	0	1	0	1	1	1	0	
3 Februari 2019	Hapsah Dg Rotang	0	1	43	Sekeloa		0	1	1	1	0	1	0	1	0	
5 Februari 2019	Jayanti Dg Soma	0	1	27	Bontomani		1	0	1	1	0	1	1	1	0	
6 Februari 2019	Auli	0	1	32	Pondang		1	0	0	1	0	1	1	1	0	
10 Februari 2019	Pelias	0	1	19	Bontomani		1	0	1	1	0	1	1	1	0	
9 Februari 2019	Dg Jinta	0	1	72	Sekeloa		1	1	1	1	1	0	0	1	0	

3. Metode *clustering* yang digunakan adalah *K-Means Clustering* dengan melakukan reduksi dimensi menggunakan algoritma PCA dan mengoptimalkan titik centroid menggunakan algoritma PSO.

## VII. Penelitian Terkait

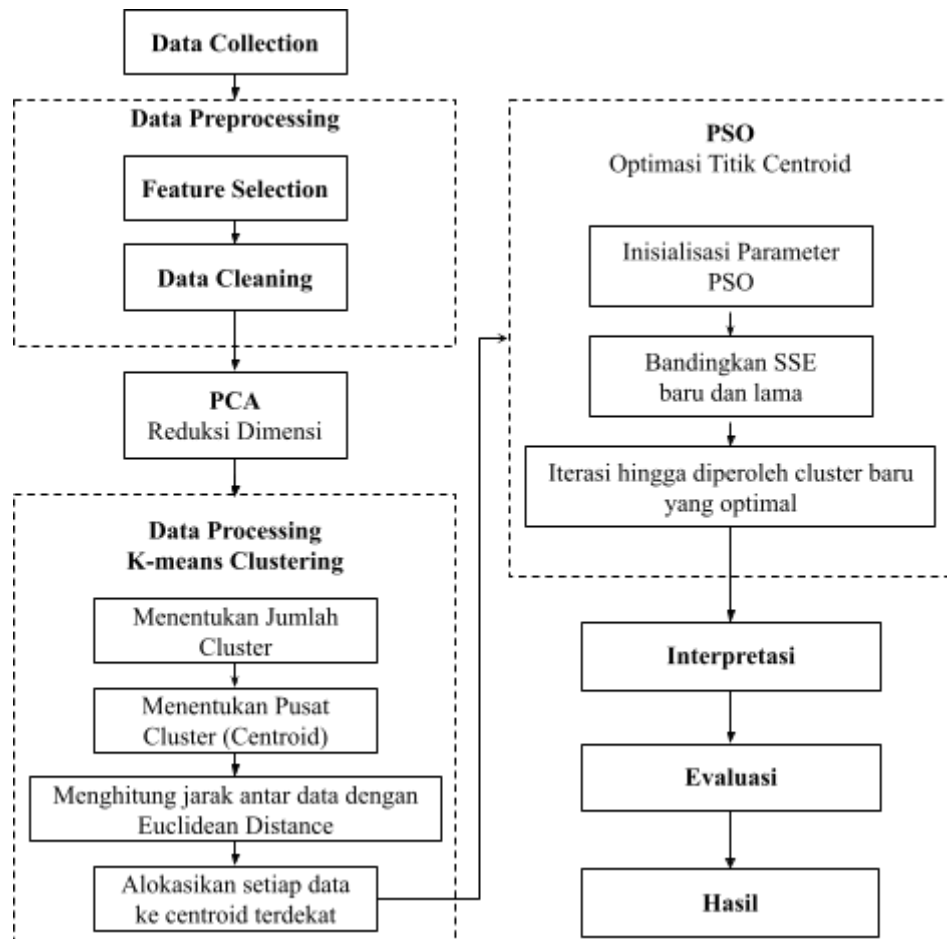
Judul	Penulis	Penerbit/Tahun	Metode	Hasil
Analisis K-Means Clustering pada Data Sepeda Motor	Rozzi Kesuma Dinata, dkk	Informatics Journal, 2020	K-Means	Hasil analisis performansi k-means dari 15 pengujian dari setiap uji coba, diperoleh nilai rata-rata Precision sebesar 76%, nilai Recall sebesar 76%, dan Accuracy sebesar 77%
Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah	Castaka Agus Sugianto, dkk	JOINT (Journal of Information Technology), 2022	K-Means, K-Medoids	Nilai Davies Bouldin k-means sebesar -0.453 sedangkan k-medoids sebesar -1.276 sehingga algoritma k-means lebih baik dari algoritma k-medoids dengan tingkat akurasi 50%.
Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA	Ikhsan Romli, dkk	Indonesian Journal of Business Intelligence, 2021	K-Means	Pengujian cluster 1 (ISPA biasa) menghasilkan nilai DBI -0.244, cluster 2 (ISPA sedang) menghasilkan nilai DBI -0.239, dan cluster 3 (ISPA berat) menghasilkan nilai DBI -0.239.

Penerapan PCA untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian di Kabupaten Bojonegoro	Dyah Hedyati, dkk	Journal Information Engineering and Educational Technology, 2021	K-Means, PCA	Evaluasi data hasil clustering menggunakan nilai DB index menunjukkan nilai paling optimal pada dataset yang direduksi menjadi 1 PC yaitu 0.4072 sedangkan dataset tanpa proses reduksi dimensi menghasilkan nilai DB Index 0.4598
Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma K-Means Clustering Berbasis Particle Swarm Optimization	Oding Herdiana, dkk	Jurnal Nuansa Informatika, 2021	K-Means, SVM, PSO	<ul style="list-style-type: none"> <li>- Hasil akurasi yang dihasilkan SVM tanpa menggunakan K-Means dan PSO adalah 50%</li> <li>- Hasil akurasi yang dihasilkan SVM dengan menggunakan K-Means dan PSO adalah 60.17%</li> </ul>
Data Mining Optimization Based on Particle Swarm Optimization for Diagnosis of Inflammatory Liver Disease	Amrin, dkk	JITE (Journal of Informatics and Telecommunicati on Engineering)	Algoritma C4.5 Naive Bayes, K-NN, PSO	Tingkat akurasi performa model C4.5 untuk diagnosa penyakit peradangan hati sebesar 70.99%, K-Nearest Neighbor sebesar 67.19%, model C4.5 dengan PSO sebesar 79.51%, dan k-NN dengan PSO sebesar 75.99%
Penerapan Metode Principal Component Analysis (PCA)	Elly Muningsih, dkk	Jurnal Komputer dan Informatika Akademi Bina Sarana Informatika	K-Means, PCA	Hasil modelling K-Means dan PCA menghasilkan nilai DBI yang lebih baik dibanding modelling

untuk Clustering Data Kunjungan Wisatawan Mancanegara di Indonesia		Yogyakarta, 2020		K-Means dengan tingkat akurasi sebesar 65%.
Spatial Selection Optimization For EEG Channel Selection Using Particle Swarm Optimization	Fadly Muhammad Kamil, dkk	e-Proceeding of Engineering, 2021	K-NN, PSO	Sistem seleksi kanal EEG dengan metode PSO menghasilkan 7 kanal aktif dengan rata-rata tingkat akurasi 100% dibanding hasil seleksi kanal EEG dengan menggunakan spatial selection dengan tingkat akurasi 50%.
Peningkatan Kinerja Algoritma K-Means Dengan Menggunakan Particle Swarm Optimization dalam Pengelompokan Data Penyediaan Akses Sanitasi dan Air Bersih	Ari Yunus Hendrawan	Jurnal Elektro Luceat, 2020	K-Means, PSO	Nilai Bouldin Index menggunakan K-Means sebesar 0.208856 sedangkan dengan menggunakan K-Means PSO menghasilkan nilai bouldin index sebesar 0.08383 dengan tingkat akurasi 70%.

## VIII. Usulan Sistem





### Penjelasan:

#### - Data Collection

Tahap awal pada penelitian ini adalah mengumpulkan data dan variabel-variabel yang diperlukan terkait demam tifoid pada masyarakat di Puskesmas Bontomarannu

#### - Data Preprocessing:

##### • Feature Selection

Tahap selanjutnya yaitu *feature selection* dimana pada tahap ini akan dipilih variabel-variabel yang akan di cluster pada dataset yang diperoleh pada tahap sebelumnya.

##### • Data Cleaning

Setelah memilih variabel-variabel apa yang akan di cluster, maka langkah selanjutnya adalah tahap data cleaning dimana pada tahap ini data akan dibersihkan seperti menghapus data duplikat, mengatasi kesalahan/*error* pada variabel, mengatasi data yang kosong, dan menghapus/mengatasi *outlier* apabila ada.

- **Reduksi Dimensi Data (Algoritma PCA)**

Selanjutnya adalah melakukan reduksi dimensi menggunakan algoritma *Principal Component Analysis* (PCA) terhadap data untuk mengurangi kesalahan pada proses klasifikasi akibat dimensi dan kompleksitas data yang tinggi sehingga mempengaruhi hasil klasifikasi. PCA akan membentuk sekumpulan dimensi baru yang kemudian diurutkan berdasarkan varian datanya

- **Data Processing (K-means Clustering)**

Setelah melakukan reduksi dimensi menggunakan algoritma *Principal Component Analysis* (PCA), maka selanjutnya akan dilakukan tahap processing data yang diawali dengan menentukan jumlah cluster yang diinginkan, setelah itu menentukan pusat cluster (centroid) dari setiap variabel di setiap cluster, dan kemudian 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.

- **Optimasi Titik Centroid (Algoritma PSO)**

Setelah melakukan clustering data, tahap berikutnya adalah melakukan pengoptimalan titik centroid menggunakan metode *Particle Swarm Optimization* (PSO) yang diawali dengan inisialisasi parameter PSO dan membandingkan nilai *Sum of Squares Error* (SSE) yang lama dan baru secara berulang hingga diperoleh hasil cluster yang optimal.

- **Interpretasi**

Setelah didapatkan hasil optimal dari proses *clustering*, kemudian hasil dari clustering akan diinterpretasi untuk mengetahui karakteristik pada cluster-cluster pada data demam tifoid pada pasien.

- **Evaluasi**

Selanjutnya setelah melakukan interpretasi, kemudian dilakukan evaluasi hasil cluster dengan Davies-bouldin index (DBI) dimana evaluasi akan dibagi 2 hasil yaitu clustering setelah menggunakan metode K-means dan clustering setelah dioptimasi dengan algoritma PCA dan PSO untuk melihat apakah hasil clustering mana yang lebih optimal

- **Hasil**

Setelah dilakukan evaluasi, maka akan diperoleh hasil dari clustering yang optimal terkait karakteristik gejala demam tifoid yang akan ditampilkan dalam bentuk peta digital dan bisa dijadikan acuan oleh pihak Puskesmas atau bahkan masyarakat sebagai informasi terkait gejala demam tifoid.

## **IX. Daftar Pustaka**

- [1] “Dinata, R. K., Safwandi, Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. Informatics Journal: Vol.5 No.1 ”
- [2] “Hediyati, D., & Suartana, I. M. (2021). Penerapan PCA untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian di Kabupaten Bojonegoro. Journal Information Engineering and Educational Technology: Vol. 05 No. 02 ”
- [3] “Hendrawan, A. Y. (2020). Peningkatan Kinerja Algoritma K-Means Dengan Menggunakan Particle Swarm Optimization dalam Pengelompokan Data Penyediaan Akses Sanitasi dan Air Bersih. Jurnal Elektro Luceat: Vol. 6 No. 2”
- [4] “Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2022). Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah. Journal of Information Technology: Vol. 02 No. 02”
- [5] “Romli, I., & Dewi, R. F. P. (2021). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA. Indonesian Journal of Business Intelligence: Vol. 04 Issue 01”
- [6] “Herdiana, O., Maulani, S., & Firdaus, E. A. (2021). Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma K-Means

Clustering Berbasis Particle Swarm Optimization. Jurnal Nuansa Informatika: Vol. 15 No. 2”

- [7] “Amrin, & Pahlevi, O. (2021). Data Mining Optimization Based on Particle Swarm Optimization for Diagnosis of Inflammatory Liver Disease. Journal of Informatics and Telecommunication: Vol. 05 No. 01”
- [8] “Elly Muningsih, E., Hasan, N., & Sulisty, G. B. (2020). Penerapan Metode Principal Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara di Indonesia. Jurnal Komputer dan Informatika Akademi Bina Sarana Informatika Yogyakarta: Vol. 08 No. 01”
- [9] “Kamil, F. M., Putra, H. F., & Ibrahim, N. (2021). Spatial Selection Optimization For EEG Channel Selection Using Particle Swarm Optimization. e-Proceeding of Engineering: Vol. 08 No. 06”
- [10] “WHO. 2018. Weekly Epidemiological Record. Geneva: WHO”