



SKRIPSI

**OPTIMASI ALGORITMA GAUSSIAN MIXTURE
MODEL MENGGUNAKAN PCA - UMAP
UNTUK KLASTERISASI TINGKAT
KESEJAHTERAAN RAKYAT**

RAFIF ILAFI WAHYU GUNAWAN
NPM 21081010093

DOSEN PEMBIMBING

Muhammad Muharrom Al Haromainy, S.Kom, M.Kom
Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

**OPTIMASI ALGORITMA *GAUSSIAN MIXTURE MODEL*
MENGUNAKAN PCA - UMAP UNTUK KLASTERISASI TINGKAT
KESEJAHTERAAN RAKYAT**

Oleh :

Rafif Ilafi Wahyu Gunawan
NPM. 21081010093

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 26 November 2025

Menyetujui

Muhammad Muharrom Al Haromainy,
S.Kom., M.Kom.

NIP. 19950601 202203 1 006



(Pembimbing I)

Achmad Junaidi, S.Kom., M.Kom

NIP. 19781110 202521 1 048



(Pembimbing II)

Eka Prakarsa Mandyartha, S.T., M.Kom

NIP. 19880525 201803 1 001



(Ketua Penguji)

Chrystia Aji Putra, S.Kom, M.T

NIP. 19861008 202121 1 001



(Anggota Penguji)

Mengetahui,

Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**OPTIMASI ALGORITMA *GAUSSIAN MIXTURE MODEL*
MENGUNAKAN PCA – UMAP UNTUK KLASTERISASI TINGKAT
KESEJAHTERAAN RAKYAT**

Oleh

Rafif Ilafi Wahyu Gunawan

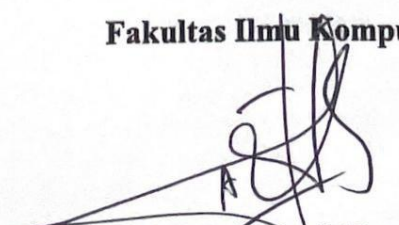
NPM. 21081010093



Menyetujui,

Koordinator Program Studi Informatika

Fakultas Ilmu Komputer


Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama : Rafif Ilafi Wahyu Gunawan
NPM : 21081010093
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa adanya paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 14 Januari 2026
Mahasiswa,



Rafif Ilafi Wahyu Gunawan
NPM. 21081010093

ABSTRAK

Nama Mahasiswa / NPM : Rafif Ilafi Wahyu Gunawan / 21081010093
Judul Skripsi : Optimasi Algoritma Gaussian Mixture Model
Menggunakan PCA – UMAP Untuk Klasterisasi
Tingkat Kesejahteraan Rakyat
Dosen Pembimbing : 1. Muhammad Muharrom Al Haromainy, S.Kom.,
M.Kom.
2. Achmad Junaidi, S.Kom., M.Kom.

Kesenjangan kesejahteraan antarwilayah masih menjadi permasalahan utama dalam pembangunan daerah, khususnya di Provinsi Jawa Timur. Perbedaan kondisi sosial, ekonomi, dan kualitas hidup menyebabkan distribusi kesejahteraan yang tidak merata. Untuk menganalisis permasalahan tersebut, penelitian ini menerapkan algoritma *Gaussian Mixture Model (GMM)* yang dioptimalkan melalui kombinasi reduksi dimensi *Principal Component Analysis (PCA)* dan *Uniform Manifold Approximation and Projection (UMAP)*. Data yang digunakan berasal dari Badan Pusat Statistik (BPS) periode 2010–2024. Hasil penelitian menunjukkan bahwa kombinasi PCA–UMAP menghasilkan nilai *Silhouette Score* tertinggi sebesar 42.36% dengan jumlah klaster optimal tiga $K = 3$, yang menggambarkan pemisahan wilayah dengan tingkat kesejahteraan tinggi dan rendah secara jelas. Pendekatan ini terbukti efektif dalam menghasilkan klaster yang stabil, representatif, dan mudah diinterpretasikan, sehingga dapat digunakan sebagai dasar dalam penyusunan kebijakan pembangunan berbasis data yang lebih tepat sasaran dan berkelanjutan.

Kata kunci : *Gaussian Mixture Model (GMM)*, PCA–UMAP, Klasterisasi, Kesejahteraan Rakyat.

ABSTRACT

Student Name / NPM : Rafif Ilafi Wahyu Gunawan / 21081010093
Thesis Title : *Optimization of the Gaussian Mixture Model Algorithm Through PCA–UMAP Integration for Clustering Public Welfare Levels*
Advisor : 1. Muhammad Muharrom Al Haromainy, S.Kom., M.Kom.
2. Achmad Junaidi, S.Kom., M.Kom

Socioeconomic inequality among regions remains a major challenge in regional development, particularly in East Java Province. Differences in social, economic, and quality-of-life conditions have led to an uneven distribution of welfare levels. To analyze this issue, this study applies the Gaussian Mixture Model (GMM) optimized through a combination of Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) for welfare clustering. The dataset used was obtained from the Central Bureau of Statistics (BPS) for the period 2010–2024. The results show that the PCA–UMAP combination achieved the highest Silhouette Score of 42.36% with an optimal cluster number of three $K = 3$, effectively distinguishing regions with high and low welfare levels. This approach proved to be efficient in producing stable, representative, and interpretable clusters, making it a potential tool to support data-driven policymaking aimed at equitable and sustainable regional development.

Keywords: *Gaussian Mixture Model (GMM), PCA–UMAP, Clustering, Public Welfare*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Optimasi Algoritma *Gaussian Mixture Model* Menggunakan *PCA* – *UMAP* Untuk Klasterisasi Tingkat Kesejahteraan Rakyat”** dapat terselesaikan dengan baik.

Penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Penulis menyampaikan apresiasi sebesar besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi selama proses penelitian ini yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan serta memotivasi kepada penulis dalam menyelesaikan skripsi. Penulis menyadari bahwa penyelesaian skripsi ini tidak lepas dari bantuan, bimbingan, serta dorongan baik secara morial, spiritual, maupun materiil dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT. Selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
4. Bapak Muhammad Muharrom Al Haromainy, S.Kom, M.Kom, selaku Dosen Pembimbing I, yang telah membimbing penulis selama pengerjaan penelitian ini.
5. Bapak Achmad Junaidi, S.Kom, M.Kom, selaku Dosen Pembimbing II, yang telah membimbing penulis selama pengerjaan penelitian ini.
6. Bapak Eka Prakarsa Mandyartha, ST, M.Kom selaku Dosen Penguji I, yang telah menguji penulis dengan memberikan perspektif baru pada penelitian ini.
7. Bapak Chrystia Aji Putra, S.Kom, M.T selaku Dosen Penguji II, yang telah menguji penulis dengan memberikan perspektif baru pada penelitian ini.

8. Bapak Firza Prima Aditiawan, S.Kom., MTI, selaku Dosen Wali, yang telah mengarahkan penulis selama perkuliahan dan penelitian ini.
9. Dosen dan Staff Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur atas segala bantuan untuk mempermudah penulis dalam penyusunan penelitian ini.
10. Ayah, Ibu, serta Kakak dan Saudara yang tidak pernah bosan untuk mendoakan dan memberikan semangat, serta dukungan materiil kepada penulis selama menempuh studi.
11. Teman-teman yang telah memberikan dukungan selama keberlangsungan masa perkuliahan dan penyusunan penelitian.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 26 November 2026

Penulis,

Rafif Ilafi Wahyu Gunawan

NPM. 21081010093

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN.....	iv
SURAT PERNYATAAN BEBAS PLAGIASI	v
ABSTRAK.....	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian	5
1.5. Batasan Penelitian.....	5
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian Terdahulu	6
2.2. Kesejahteraan Rakyat	8
2.3. Pemrosesan Data (<i>Preprocessing</i>)	10
2.4. Reduksi Dimensi (<i>Dimensionality Reduction</i>).....	11
2.5. <i>Uniform Manifold Approximation and Projection</i> (UMAP)	12
2.6. <i>Bayesian Information Criterion</i> (BIC)	14
2.7. Klastering (Clustering)	15
2.8. Gaussian Mixture Model (GMM).....	15
2.9. <i>Expectatio</i> -Maximization (EM).....	15
2.10. Pelabelan Kluster	16
2.11. Silhoutte Score.....	17
BAB III DESAIN DAN IMPLEMENTASI SISEM	18
3.1. Tahapan Penelitian	18
3.2. Pengumpulan Dataset	18

3.3. Pemrosesan Data (<i>Data Preprocessing</i>).....	21
3.5.1. Pembersihan Data (<i>Data Cleaning</i>)	21
3.5.2. Normalisasi Data (<i>Data Normalization</i>)	24
3.5.3. Reduksi Dimensi (<i>Dimensionality Reduction</i>).....	27
3.5.4. <i>Uniform Manifold Approximation and Projection</i> (UMAP)	34
3.4. Pembuatan Model	41
3.4.1. Gaussian Mixture Model (GMM).....	41
3.4.2. <i>Bayesian Information Criterion</i> (BIC)	56
3.4.3. Pelabelan Hasil Kluster.....	57
3.5. Evaluasi Hasil	60
3.6. Skenario Pengujian	62
3.7. Analisis Hasil.....	64
BAB IV	65
PENGUJIAN DAN ANALISA	65
4.1. Pengumpulan Data.....	65
4.2. <i>Preprocessing</i> Data.....	65
4.2.1 Pembersihan Data (<i>Data Cleaning</i>)	66
4.2.2 Normalisasi Data (<i>Data Normalization</i>)	68
4.2.3 Reduksi Dimensi (<i>Dimensionalty Reduction</i>).....	72
4.2.4 <i>Uniform Manifold Approximation and Projection</i> (UMAP)	80
4.3. Pembuatan Model	85
4.3.1. <i>Gaussian Mixture Model</i> (GMM).....	86
4.3.2. <i>Bayessian Information Criterion</i> (BIC)	89
4.3.3. Pelabelan Kluster	93
4.4. Skenario Pengujian	98
4.5. Evaluasi Hasil.....	102
4.5.1. Evaluasi Model <i>Bayessian Information Criterion</i> (BIC).....	102
4.5.2. Evaluasi Model <i>Silhouette Score</i>	103
4.6. Analisis Hasil.....	104
4.6.1. Analisis Karakteristik Kluster.....	104
4.6.2. Distribusi Wilayah Per Kluster	105
4.6.3. Interpretasi dan Implikasi Hasil.....	106

BAB V	108
PENUTUP	108
DAFTAR PUSTAKA	109
LAMPIRAN	113

DAFTAR GAMBAR

Gambar 3. 1 Metodologi Penelitian.....	18
Gambar 3. 2 Tahapan Pemrosesan Data.....	21
Gambar 3. 3 Tahapan Pembersihan Data	22
Gambar 3. 4 Tahapan Normalisasi Data (<i>Min-Max</i>).....	24
Gambar 3. 5 Tahapan <i>Principal Component Analysis</i> (PCA).....	27
Gambar 3. 6 Tahapan UMAP.....	34
Gambar 3. 7 Tahapan Pembuatan Model	41
Gambar 3. 8 Tahapan <i>Gaussian Mixture Model</i> (GMM).....	42
Gambar 3. 9 Tahapan <i>Bayessian Information Criterion</i> (BIC).....	56
Gambar 4. 1 Dataset Kesejahteraan Rakyat	65
Gambar 4. 2 Cek <i>Missing Value</i>	67
Gambar 4. 3 Data <i>Missing value</i>	67
Gambar 4. 4 Hasil Data <i>Cleaning</i>	68
Gambar 4. 5 Nilai <i>Min Max</i> Setiap Indikator	69
Gambar 4. 6 Hasil Normalisasi.....	71
Gambar 4. 7 Hasil Reduksi Dimensi Data Semua Indikator	74
Gambar 4. 8 Varinasi Kumulatif	76
Gambar 4. 9 Visualisasi Varinasi Kumlatif $\leq 85\%$	77
Gambar 4. 10 <i>Factor Loading</i>	78
Gambar 4. 11 Hasil Komponen Utama	79
Gambar 4. 12 Visualisasi Proyeksi Data UMAP.....	83
Gambar 4. 13 Hasil UMAP	85
Gambar 4. 14 <i>Grid Search</i> 1.....	88
Gambar 4. 15 <i>Grid Search</i> 2.....	89
Gambar 4. 16 evaluasi <i>Silhouette</i> terbaik berdasarkan BIC	90
Gambar 4. 17 Model Terbaik	92
Gambar 4. 18 Contoh Pelabelan Klaster	94
Gambar 4. 19 Hasil Penamaan Klaster.....	96
Gambar 4. 20 Hasil Distribusi Klaster.....	97
Gambar 4. 21 Hasil Evaluasi BIC Terhadap Uji Skenario	103
Gambar 4. 22 Evaluasi K Terbaik dari Uji Skenario.....	104

Gambar 4. 23 Analisi Karakteristik Klaster	105
Gambar 4. 24 Hasil Distribusi Klaser Berdasarkan Konfigurasi Terbaik	106
Gambar 4. 25 Hasil Interpretasi Tahun 2024	107

DAFTAR TABEL

Tabel 3. 1 Rincian Variabel Pada Dataset Kesejahteraan Rakyat	19
Tabel 3. 2 Contoh Sampel Dataset Kesejahteraan Rakyat.....	20
Tabel 3. 3 Hasil Data <i>Cleaning</i>	23
Tabel 3. 4 Hasil Data <i>Normalizaion</i>	25
Tabel 3. 5 Hasil Deviasi	27
Tabel 3. 6 Tabel Proyeksi Baru	32
Tabel 3. 7 Hasil Perhitungan Jarak Antar Titik Tetangga 1	35
Tabel 3. 8 Hasil Perhitungan Jarak Antar Titik Tetangga 2	36
Tabel 3. 9 Pengurutan Berdasarkan Jarak Dari Terkecil ke Besar	37
Tabel 3. 10 Hasil Optimasi <i>Embedding</i>	40
Tabel 3. 11 Hasil Keseluruhan <i>E-Step</i>	45
Tabel 3. 12 Hasil <i>Log-Likelihood</i>	49
Tabel 3. 13 Hasil Probabilitas (E-Step) Iterasi 2.....	51
Tabel 3. 14 Hasil Log-Likelihood Iterasi 2	53
Tabel 3. 15 Label Hasil Klaster	57
Tabel 3. 16 Hasil Pelabelan Status.....	60
Tabel 3. 17 Hasil <i>Silhouutte Score</i>	62
Tabel 3. 18 Skenario Pengujian.....	63
Tabel 4. 1 Skenario Pengujian <i>NoReduksi</i>	98
Tabel 4. 2 Skeanrio Pengujian PCA	99
Tabel 4. 3 Skenario Pengujian PCA-UMAP	100
Tabel 4. 4 Hasil Terbaik Dari Masing-Masing Uji Skenario.....	102