



SKRIPSI

**KLASTERISASI KABUPATEN/KOTA DI JAWA
TIMUR BERDASARKAN FAKTOR RISIKO
PENULARAN TUBERKULOSIS MENGGUNAKAN
FUZZY GUSTAFSON KESSEL DENGAN
VALIDITAS MODIFIED PARTITION
*COEFFICIENT (MPC)***

NAILA MUGHNIFA QALBI
NPM 22083010022

DOSEN PEMBIMBING

Aviolla Terza Damaliana, S.Si.,M.Stat.
Dr.Ir.Mohammad Idhom, S.P.,S.Kom.,M.T.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN


**KLASTERISASI KABUPATEN/KOTA DI JAWA TIMUR
BERDASARKAN FAKTOR RISIKO PENULARAN TUBERKULOSIS
MENGUNAKAN *FUZZY GUSTAFSON KESSEL* DENGAN VALIDITAS
*MODIFIED PARTITION COEFFICIENT (MPC)***

Oleh:
NAILA MUGHNIFA QALBI
NPM. 22083010022

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 11 Juni 2026.

Menyetujui,

Aviolla Terza Damaliana, S.Si., M.Stat
NIP. 19940802 202203 2 015


.....

(Pembimbing I)

Dr.Ir.Mohammad Idhom, S.P., S.kom., M.T.
NIP. 19830310 202121 1 006


.....

(Pembimbing II)

Amri Muhaimin, S.Stat., M.Stat., MS.
NIP. 19950723 202406 1 002


.....


(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom
NIP. 19990606 202406 1 001


.....

(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer

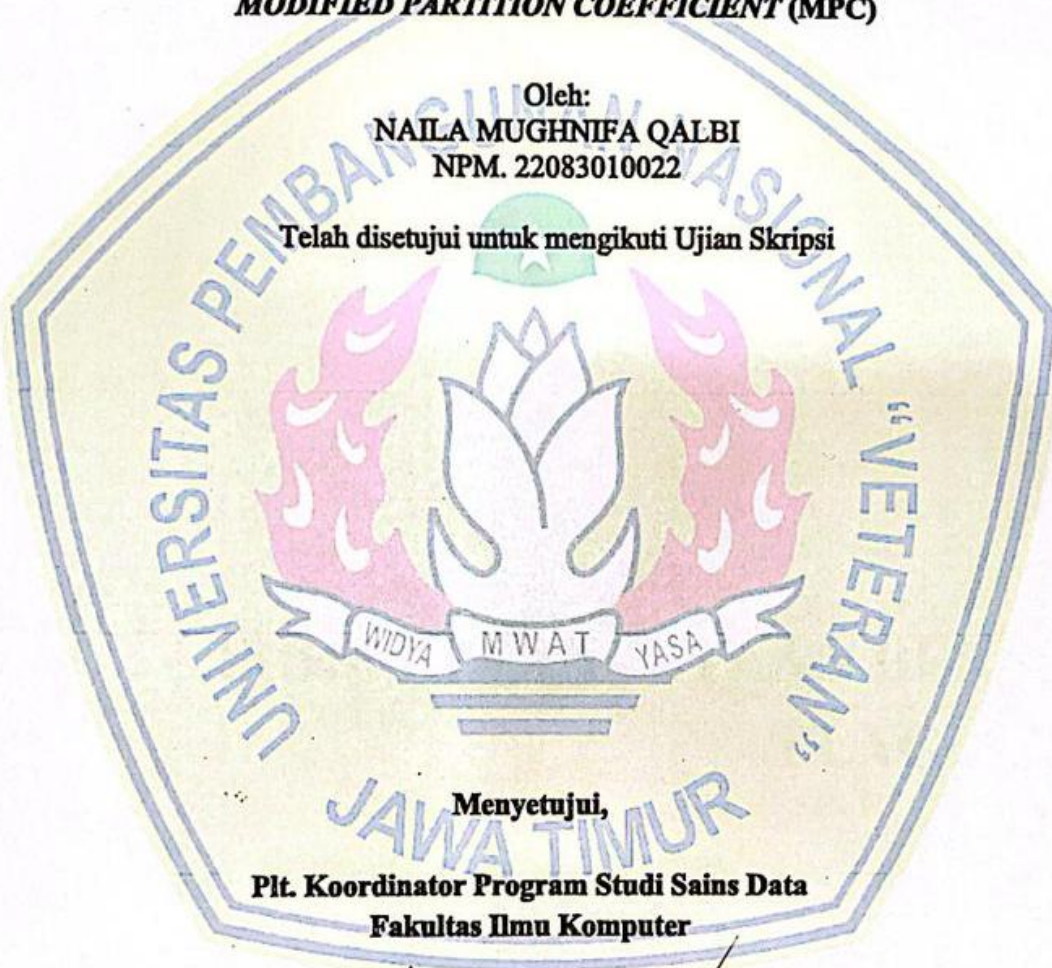

Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 196811261994032001

LEMBAR PERSETUJUAN

**KLASTERISASI KABUPATEN/KOTA DI JAWA TIMUR
BERDASARKAN FAKTOR RISIKO PENULARAN TUBERKULOSIS
MENGUNAKAN *FUZZY GUSTAFSON KESSEL* DENGAN VALIDITAS
*MODIFIED PARTITION COEFFICIENT (MPC)***

Oleh:
NAILA MUGHNIFA QALBI
NPM. 22083010022

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,
Plt. Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. I Gede Susrama Mas Divasa, S.T., M.T.
NIP. 19700619 202121 1 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Naila Mughnifa Qalbi
NPM : 22083010022
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : 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 di kemudian 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 ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 11 Juni 2026
Yang Membuat Pernyataan,



Naila Mughnifa Qalbi
NPM. 22083010022

ABSTRAK

Nama Mahasiswa / NPM : Naila Mughnifa Qalbi / 22083010022
Judul Skripsi : Klasterisasi Kabupaten/Kota Di Jawa Timur Berdasarkan Faktor Risiko Penularan Tuberkulosis Menggunakan *Fuzzy Gustafson Kessel* Dengan Validitas *Modified Partition Coefficient* (MPC)
Dosen Pembimbing : 1. Aviolla Terza Damaliana,S.Si.,M.Stat.
2. Dr.Ir.Mohammad Idhom,S.P.,S.kom.,M.T.

Tuberkulosis (TB) merupakan penyakit menular di Indonesia yang terus meningkat dan menjadi tantangan bagi pemerintah dalam mengurangi angka kejadian penyakit tersebut. Pada tahun 2024, Indonesia menempati peringkat kedua di dunia dengan 1.090.000 kasus TB. Provinsi Jawa Timur termasuk daerah dengan jumlah kasus tertinggi dan menempati peringkat ketiga dengan total 88.733 kasus. Namun, tingkat deteksi kasus di daerah tersebut mengalami penurunan. Kondisi ini menunjukkan perlunya strategi pengendalian penyakit yang lebih terarah. Studi ini bertujuan untuk mengelompokkan 38 kabupaten/kota di Provinsi Jawa Timur berdasarkan faktor risiko TB, yaitu jumlah penderita HIV, penderita diabetes melitus, balita dengan gizi buruk, kepadatan penduduk, jumlah penduduk laki-laki, jumlah penduduk miskin, dan jumlah perokok berusia 15–64 tahun. Pengelompokan dilakukan menggunakan metode *Fuzzy Gustafson Kessel* (FGK) dengan pendekatan jarak Mahalanobis, untuk membentuk klaster berdasarkan karakteristik data dengan pola distribusi yang berbeda. Berdasarkan hasil pengujian beberapa klaster, jumlah klaster optimal adalah 3 klaster dengan nilai *Modified Partition Coefficient* (MPC) sebesar 0,7613, dan *Entropi Partisi* (PE) sebesar 0,3198. Hasil pengelompokan menunjukkan bahwa klaster 2 dikategorikan sebagai prioritas tinggi dengan 15 wilayah sebagai anggota, klaster 1 dikategorikan sebagai prioritas menengah dengan 11 wilayah sebagai anggota. Sementara itu, klaster 3 dikategorikan sebagai prioritas rendah dengan 12 wilayah. Selain itu, metode *Fuzzy Gustafson Kessel* berhasil diimplementasikan dalam bentuk GUI berbasis web, sehingga proses pengunggahan data, eksplorasi data, pengelompokan, interpretasi hasil, dan visualisasi peta dapat dilakukan dengan cara yang lebih terstruktur.

Kata Kunci: *Tuberkulosis, Fuzzy Gustafson Kessel, Modified Partition Coefficient, Klasterisasi Wilayah, Risiko penularan TBC*

ABSTRACT

Student Name / NPM : Naila Mughnifa Qalbi / 22083010022
Title : *Clustering Of Regencies/Cities In East Java Based On Tuberculosis Transmission Risk Factors Using Fuzzy Gustafson Kessel With Modified Partition Coefficient (Mpc) Validity*
Advisors : 1. Aviolla Terza Damaliana,S.Si.,M.Stat.
2. Dr.Ir.Mohammad Idhom,S.P.,S.kom.,M.T.

Tuberculosis (TB) is an infectious disease in Indonesia that continues to increase and poses a challenge for the government in reducing the incidence of the disease. In 2024, Indonesia ranked second in the world with 1,090,000 tuberculosis cases. East Java Province is among the regions with the highest number of cases and ranked third with a total of 88,733 cases. However, the case detection rate in the region has decreased. This condition indicates the need for a more targeted disease control strategy. This study aims to group 38 districts/cities in East Java Province based on tuberculosis risk factors, namely the number of HIV sufferers, diabetes mellitus sufferers, toddlers with malnutrition, population density, number of male residents, number of poor residents, and number of smokers aged 15–64 years. Clustering was carried out using the Fuzzy Gustafson-Kessel (FGK) method with the Mahalanobis distance approach, to form clusters based on data characteristics with different distribution patterns. Based on the results of testing several clusters, the optimal number of clusters was 3 clusters with a Modified Partition Coefficient (MPC) value of 0.7613, and a Partition Entropy (PE) of 0.3198. The clustering results showed that cluster 2 was categorized as high priority with 15 regions as members, cluster 1 was categorized as medium priority with 11 regions as members. Meanwhile, cluster 3 was categorized as low priority with 12 regions. In addition, the Fuzzy Gustafson Kessel method was successfully implemented in the form of a web-based Graphical User Interface (GUI), so that the process of data upload, data exploration, clustering, interpretation of results, and map visualization could be done in a more structured and interactive manner.

Keywords: *Tuberculosis, Fuzzy Gustafson Kessel, Modified Partition Coefficient, Regional Clustering, Risk of TB Transmission.*

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
DAFTAR NOTASI	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	8
1.3. Batasan Masalah	8
1.4. Tujuan Penelitian	9
1.5. Manfaat Penelitian	9
BAB II TINJAUAN PUSTAKA	11
2.1. Penelitian Terdahulu	11
2.2. Kerangka Teori	22
2.2.1. Tuberkulosis	22
2.2.2. Analisis Klustering	23
2.2.3. Standarisasi Data	24
2.2.4. Teori Fuzzy	25
2.2.5. Fuzzy C-Means	26
2.2.6. Jarak Mahalanobis.....	30
2.2.7. Fuzzy Gustafson kessel	30
2.2.8. Modified Partition Coefficient (MPC)	36

2.2.9.	PE (<i>Partititon Entropyt</i>).....	37
2.2.10.	GUI	37
BAB III	METODOLOGI PENELITIAN.....	39
3.1.	Variabel Penelitian dan Sumber Data	39
3.2.	Langkah Analisis	43
3.3.	Desain Sistem	47
BAB IV	HASIL DAN PEMBAHASAN	51
4.1.	Pengumpulan Data	51
4.2.	<i>Preprocessing</i> Data	52
4.3.	Implementasi <i>Fuzzy Gustafson Kessel</i> Manual	59
4.4.	Hasil Uji Beberapa Parameter	77
4.5.	Perbandingan <i>Fuzzy C-Means</i> dan <i>Fuzzy Gustafson Kessel</i>	88
4.6.	Profilisasi Hasil Kluster	90
4.7.	GUI.....	94
4.7.1.	Halaman <i>Home</i>	94
4.7.2.	Halaman Eksplorasi Data.....	95
4.7.3.	Halaman Interpretasi Kluster	96
4.7.4.	Halaman Visualisasi Peta.....	97
5.1.	Kesimpulan.....	99
5.2.	Saran Pengembangan	100
DAFTAR PUSTAKA		103
LAMPIRAN		111

DAFTAR GAMBAR

Gambar 2.1 Flowchart Alur <i>Fuzzy C Means</i>	29
Gambar 2.2 Flowchart <i>Fuzzy Gustafson Kessel</i>	32
Gambar 3.1 <i>Flowchart</i> Alur Penelitian	44
Gambar 3.2 Desain Sistem	48
Gambar 3.3 Diagram Alir User Sistem	50
Gambar 4.1 Peta Persebaran Tuberkulosis di Jawa Timur.....	56
Gambar 4.2 Plot Visualisasi Hasil Klaster	87
Gambar 4.3 Visualisasi Peta.....	93
Gambar 4.4 Halaman <i>Home</i>	94
Gambar 4.5 Halaman Eksplorasi Data	95
Gambar 4.6 Halaman Interpretasi Klaster	97
Gambar 4.7 Halaman Visualisasi Peta	98

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	11
Tabel 3.1 Variabel Penelitian	39
Tabel 3.2 Kerangka Dataset Penelitian	43
Tabel 3.3 Perbandingan Nilai Parameter.....	46
Tabel 4.1 Dataset Penelitian.....	51
Tabel 4.2 Statistika Deskriptif.....	53
Tabel 4.3 Hasil Standarisasi Data	58
Tabel 4.4 Nilai Derajat Keanggotaan Awal	61
Tabel 4.5 Normalisasi Nilai Derajat Keanggotaan.....	62
Tabel 4.6 Hasil Pusat Klaster	64
Tabel 4.7 Hasil Nilai R_{ci}	66
Tabel 4.8 Nilai Tranpose R_{ci} (R_{ciT})	66
Tabel 4.9 Hasil Nilai $R_{ci} *$	67
Tabel 4.10 Hasil Matriks Kovarians Klaster 2 (F_{c2})	68
Tabel 4.11. Hasil Jarak Klaster	72
Tabel 4.12 Nilai Iterasi Fungsi Objektif.....	75
Tabel 4.13 Nilai Keanggotaan Akhir	76
Tabel 4.14 Hasil Nilai MPC	77
Tabel 4.15 Hasil Evaluasi Klaster	81
Tabel 4.16 Nilai Derajat Keanggotaan Akhir.....	82
Tabel 4.17 <i>Centroid</i> Klaster	82
Tabel 4.18 Matriks Kovarians <i>Fuzzy</i> Klaster 1	83
Tabel 4.19 Matriks Kovarians <i>Fuzzy</i> klaster 2	84
Tabel 4.20 Matriks Kovarians <i>Fuzzy</i> Klaster 3	85
Tabel 4.21 Hasil Anggota Klaster	88
Tabel 4.22 Perbandingan FCM dan FGK.....	89
Tabel 4.23 Rata-Rata Profil Setiap Klaster	90

DAFTAR LAMPIRAN

Lampiran 1. Dataset	111
Lampiran 2. Kode Program.....	112
Lampiran 3. GUI	113
Lampiran 4. LOA Jurnal	114

DAFTAR NOTASI

n	:	Jumlah data
p	:	Jumlah variabel
r_{ij}	:	Koefisien korelasi variabel i dan j
a_{ij}	:	Koefisien korelasi parsial antara variabel i dan j
R_j^2	:	Koefisien determinasi
x_i	:	Data ke - i
s	:	Standar deviasi dari data
a_{ic}	:	Inisialisasi awal
x	:	Rata-rata data
m	:	Pangkat pembobot
c	:	Jumlah kluster
ε	:	Error terkecil
P_0	:	Inisialisasi awal fungsi objektif
U_0	:	Inisialisasi awal dari matriks partisi U_0
c_i	:	Indeks kluster
$a_{ic_i}(x_i, v_{c_i})$:	Jarak data ke pusat kluster
$a_{ic_i}(v_i, v_{c_i})$:	Fungsi gaussian satu dimensi
b_{ic_i}	:	Rata-rata jarak data ke i titik dalam kluster terdekat lainnya
a_{ic_i}	:	Derajat keanggotaan tertinggi data ke i pada kluster ke c_i
F_{c_i}	:	Matriks kovarian kluster
$D_{ic_i}^2$:	Jarak mahalanobis
A_{c_i}	:	Fungsi Jarak Objektif
$S_{i,q}$:	Jarak mahalanobis antara data ke i dalam kluster ke kluster lain (q)

R_{c_i}	:	Selisih data ke- i dengan pusat kluster pada masing–masing variabel
$R_{c_i}^*$:	Matriks yang berasal dari perkalian $a_{ic_i}^m$
$D_{c_i}^2$:	Jarak mahalanobis yang berasal dari variabel ke- j terhadap objek ke- i terhadap pusat kluster c .
$R_{c_i}^T$:	Hasil matriks transpose dari R_{c_i}
α	:	Bobot koefisien ($\alpha \geq 0$) dan nilai default yaitu 1
$(x_i - \bar{x})^T$:	Transpose matriks selisih data ke i dengan nilai rata-rata
$a_{ic_i}^2$:	Derajat keanggotaan dengan pangkat pembobot pada data ke i dalam kluster ke- c_i
$S_{ic_i}^2$:	Jarak mahalanobis dari diagonal utama $D_{ic_i}^2$
r_{xy}	:	Korelasi antara variabel x dan y
v_{kj}	:	Pusat kluster ke- k
μ_{ik}^m	:	Derajat keanggotaan data i ke- k
x_{ij}	:	Data input variabel i ke- j
D_{ik}^2	:	Nilai kuadrat jarak Euclidean data ke- i dan pusat kluster ke- k
P_t	:	Total fungsi objektif
S_{iq}	:	Jarak objek ke- I terhadap kluster ke- q