

PENERAPAN METODE *EXPECTATION MAXIMIZATION* – *GAUSSIAN MIXTURE MODEL* (EM-GMM) PADA SEGMENTASI CITRA DIGITAL

(Skripsi)

Oleh

MAULANA YUSUF



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRACT

APPLICATION OF EXPECTATION MAXIMIZATION – GAUSSIAN MIXTURE MODEL (EM-GMM) IN DIGITAL IMAGE SEGMENTATION

By

MAULANA YUSUF

One application of cluster analysis is digital image segmentation. Image segmentation is a process used to partition an image into simpler regions. When applying cluster analysis for image segmentation, there is a problem where the intensity of the existing pixels will attract other similar pixels into the same cluster. This problem is very rare when applying the soft clustering method because the pixels are equipped with a degree of membership. This research examines the application of the Gaussian Mixture Model, which is a soft clustering method that uses the Gaussian distribution as a component of the mixture model. The parameters of the GMM will be iterated using the Expectation-Maximization algorithm to obtain the maximum likelihood. The selection of the best number of clusters is determined based on the values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The results obtained indicate the number of clusters 7 is the optimal number of clusters.

Keywords: Cluster Analysis, Image Segmentation, Expectation-Maximization, Gaussian Mixture Model, Akaike Information Criterion, Bayesian Information Criterion.

ABSTRAK

PENERAPAN METODE *EXPECTATION MAXIMIZATION – GAUSSIAN MIXTURE MODEL* (EM-GMM) PADA SEGMENTASI CITRA DIGITAL

Oleh

MAULANA YUSUF

Salah satu aplikasi analisis cluster adalah segmentasi citra digital. Segmentasi citra merupakan proses yang digunakan untuk mempartisi citra menjadi daerah yang lebih sederhana. Saat menerapkan analisis kluster untuk segmentasi citra, terjadi permasalahan dimana intensitas piksel yang ada akan menarik piksel lainnya yang serupa ke dalam suatu kluster yang sama. Permasalahan tersebut sangat jarang terjadi jika menerapkan metode soft clustering karena pikselnya dilengkapi dengan derajat keanggotaan. Pada penelitian ini dikaji penerapan Gaussian Mixture Model yang merupakan salah satu metode soft clustering yang menggunakan distribusi Gaussian sebagai komponen dari mixture model. Parameter dari GMM akan diiterasi menggunakan algoritma Expectation-Maximization untuk mendapatkan maximum likelihood. Pemilihan jumlah kluster terbaik ditentukan berdasarkan nilai Akaike's Information Criterion (AIC) dan Bayesian Information Criterion (BIC). Hasil yang diperoleh menunjukkan jumlah kluster 7 merupakan jumlah kluster yang optimal.

Kata Kunci: Analisis Kluster, Segmentasi Citra, *Expectation-Maximization*, *Gaussian Mixture Model*, *Akaike Information Criterion*, *Bayesian Information Criterion*.

PENERAPAN METODE *EXPECTATION MAXIMIZATION* – *GAUSSIAN MIXTURE MODEL* (EM-GMM) PADA SEGMENTASI CITRA DIGITAL

Oleh

MAULANA YUSUF

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

Judul Skripsi

**: PENERAPAN METODE *EXPECTATION
MAXIMIZATION – GAUSSIAN MIXTURE
MODEL (EM-GMM)* PADA SEGMENTASI
CITRA DIGITAL**

Nama Mahasiswa

: Maulana Yusuf

Nomor Pokok Mahasiswa

: 1817031090

Program Studi

: Matematika

Fakultas

: Matematika dan Ilmu Pengetahuan Alam



1. Komisi Pembimbing

Drs. Eri Setiawan, M.Si.
NIP 195811011988031002

Dr. Muslim Ansori, S.Si., M.Si.
NIP 197202271998021001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua

: **Drs. Eri Setiawan, M.Si.**

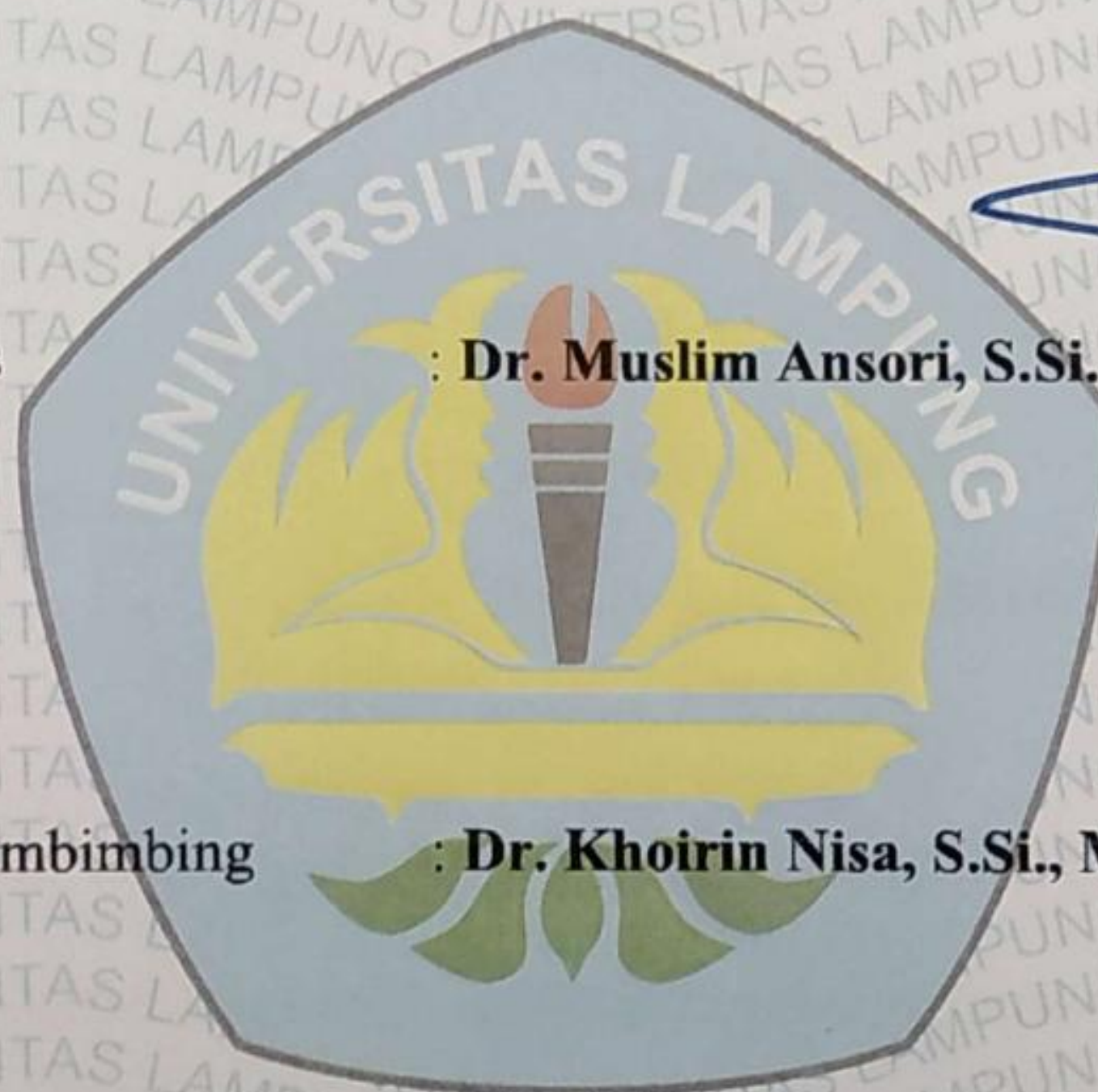
Sekretaris

: **Dr. Muslim Ansori, S.Si., M.Si.**

Penguji

Bukan Pembimbing

: **Dr. Khoirin Nisa, S.Si., M.Si.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Surtpto Dwi Yuwono, M.T.

NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 13 Juli 2022

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : **Maulana Yusuf**

Nomor Pokok Mahasiswa : **1817031090**

Jurusan : **Matematika**

Judul Skripsi : **PENERAPAN METODE *EXPECTATION*
MAXIMIZATION – *GAUSSIAN MIXTURE*
MODEL (EM-GMM) PADA SEGMENTASI
CITRA DIGITAL**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri.

Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, Juli 2022
Yang Menyatakan,



Maulana Yusuf

RIWAYAT HIDUP

Penulis memiliki nama lengkap Maulana Yusuf yang lahir di Bandar Lampung pada tanggal 24 Agustus 2000. Penulis merupakan anak bungsu dari dua bersaudara yang terlahir dari pasangan Bapak Adiluis dan Ibu Yurnita.

Penulis menyelesaikan pendidikan Sekolah Dasar di SD Muhammadiyah Pringsewu pada tahun 2012, pendidikan Sekolah Menengah Pertama di SMP Negeri 23 Bandar Lampung yang diselesaikan pada tahun 2015, dan pendidikan Sekolah Menengah Atas di SMA Negeri 7 Bandar Lampung yang diselesaikan pada tahun 2018.

Penulis melanjutkan pendidikan Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung pada tahun 2018 melalui jalur SBMPTN. Sebagai bentuk penerapan ilmu perkuliahan pada tahun 2020 dan 2021, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Kelurahan Beringin Jaya Kecamatan Kemiling Kota Bandar Lampung dan Kerja Praktik (KP) di Otoritas Jasa Keuangan.

Selama menjadi mahasiswa penulis aktif mengikuti organisasi sebagai Anggota Multimedia pada periode 2019 – 2020 di Unit Kegiatan Mahasiswa Fakultas (UKMF) Natural FMIPA Universitas Lampung. Penulis juga mengikuti kepanitiaan Dies Natalis Jurusan Matematika (DINAMIKA) sebagai Koordinator Olimpiade Matematika pada tahun 2019 – 2020. Penulis juga mengikuti salah satu program Kampus Merdeka Kementrian Pendidikan dan Kebudayaan yaitu Studi Independen Generasi Gigih – *Data Analyst Track* di Yayasan Anak Bangsa Bisa (YABB) periode Februari 2022 – Juli 2022.

KATA INSPIRASI

“Dan hanya kepada Allah hendaknya kamu berharap”

(Q.S. Al Insyirah: 8)

“Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui”

(Q.S. Al Baqarah: 216)

“Sebaik-baik manusia adalah yang paling bermanfaat bagi orang lain”

(HR. Ahmad, Tharbani, dan Daruqtni)

“Keberhasilan bukanlah milik orang yang pintar. Keberhasilan adalah kepunyaan mereka yang senantiasa berusaha”

(B.J. Habibie)

PERSEMBAHAN

Alhamdulillahillobb'il alamin,

Puji dan syukur kita haturkan kepada Allah Subhanahu Wata'ala karena atas limpahan nikmat dan karunia-Nya yang diberikan, shalawat serta salam selalu tercurah kepada Nabi Muhammad Shallallahu 'Alaihi Wassallam yang telah menuntun manusia ke jalan yang terang benderang.

Kupersembahkan karya yang sederhana ini kepada

Ayah, Ibu dan Keluargaku

Terima kasih kepada Ayah, Ibu dan Keluargaku yang selalu mendoakan kesuksesanku, memberi semangat, nasihat, dukungan serta kasih sayang yang tiada henti.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sangat berjasa dan tidak lelah memberikan arahan serta masukan sehingga tersusunlah skripsi ini.

Sahabat-sahabatku

Terima kasih kepada para sahabatku yang selalu memberikan semangat, doa, dan motivasi, serta kenangan indah selama ini.

Almamater dan Negeriku

SANWACANA

Puji dan syukur kehadiran Allah SWT atas segala rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini. Shalawat serta salam tak lupa penulis haturkan kepada Nabi Muhammad SAW atas dakwah dan risalah Beliau, sehingga kita dapat berada di zaman yang penuh ilmu. Skripsi dengan Judul “Penerapan Metode *Expectation Maximization – Gaussian Mixture Model* (EM-GMM) pada Segmentasi Citra Digital” adalah salah satu syarat untuk memperoleh gelar Sarjana Matematika di Universitas Lampung.

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena dukungan, bimbingan, saran, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Drs. Eri Setiawan, M.Si. selaku Pembimbing Akademik dan selaku Pembimbing I yang selalu bersedia memberikan arahan, bimbingan, saran, serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Dr. Muslim Ansori, S.Si., M.Si. selaku Pembimbing II yang telah memberikan arahan, bimbingan dan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si. selaku Penguji yang telah bersedia memberikan kritik dan saran serta evaluasi kepada penulis dalam penyusunan skripsi.
4. Bapak Dr. Eng. Suropto Dwi Yuwono, M.T. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Seluruh dosen, staf dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah memberikan ilmu dan bantuan kepada penulis.

6. Orang tua tercinta serta seluruh keluarga besar yang selalu memberikan doa, dukungan, dan motivasi kepada penulis.
7. Teman-teman selama di UKMF Natural, khususnya Andra, Aniisah, Ferzy, dan Sofalina yang telah memberikan perhatian dan kritik positif kepada penulis agar terus maju.
8. Teman-teman selama kepanitiaan DINAMIKA XXI, khususnya Riska, Wina, dan Wiranto yang telah membantu mengingatkan dan memotivasi penulis untuk menyelesaikan penulisan skripsi ini.
9. Samuel, Aulia, Nurfadilah, Devi, Shavira, dan Rika yang telah kebersamaan penulis selama bimbingan skripsi bersama Pak Eri.
10. Shabrina dan Virda yang telah membantu mengingatkan dan memotivasi penulis untuk menyelesaikan penulisan skripsi ini.
11. Dika, Leo, Nauval, Salsa, Prima, dan Nabila selaku teman SMA yang telah memberikan doa dan dukungan kepada penulis agar terus maju.
12. Seluruh pihak yang telah membantu penulis yang tidak dapat disebutkan satu persatu.

Penulis juga menyadari bahwa dalam penulisan skripsi ini masih banyak terdapat kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun guna penelitian selanjutnya agar lebih baik.

Bandar Lampung, Juli 2022
Penulis,

Maulana Yusuf

DAFTAR ISI

Halaman

DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	4
2.1 Analisis Kluster	4
2.2 Model Campuran	5
2.3 <i>Gaussian Mixture Model (GMM)</i>	7
2.4 Algoritma <i>Expectation-Maximization (EM)</i>	8
2.5 Pemilihan Model	11
2.5.1 <i>Akaike's Information Criterion (AIC)</i>	11
2.5.2 <i>Bayesian Information Criterion (BIC)</i>	12
2.6 Segmentasi Citra Digital	12
III. METODOLOGI PENELITIAN	14
3.1 Waktu dan Tempat Penelitian	14
3.2 Data Penelitian	14
3.3 Metode Penelitian	14

IV. HASIL DAN PEMBAHASAN	17
4.1 Persiapan dan Eksplorasi Data	17
4.1.1 Sebaran Distribusi Citra	18
4.2 Segmentasi Citra dengan Metode <i>Gaussian Mixture Model</i> (GMM).....	19
4.3 Evaluasi Hasil	22
 V. KESIMPULAN	 26
5.1 Kesimpulan	26

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel	Halaman
1. Citra RGB Setelah Dikonversi menjadi <i>Dataframe</i>	18
2. Iterasi dan Jumlah Parameter di Setiap Jumlah Klaster	20
3. Perbandingan nilai <i>Akaike Information Criterion</i> (AIC) dengan <i>Bayesian Information Criterion</i> (BIC) untuk Metode GMM	23
4. Hasil Estimasi Parameter <i>Gaussian Mixture Model</i> 7 Klaster.....	24

DAFTAR GAMBAR

Gambar	Halaman
1. Fungsi Densitas dari <i>Mixture Model</i>	5
2. Ilustrasi Citra RGB	13
3. Diagram Alir Penelitian	16
4. Citra Uji	17
5. Grafik Densitas Kernel	18
6. Fungsi Log-likelihood untuk Jumlah Klaster 2 hingga 10.....	20
7. Perbandingan Hasil Segmentasi Citra dengan Metode GMM untuk Jumlah Klaster 2 hingga 10	21
8. Perbandingan Nilai <i>Akaike Information Criterion</i> (AIC) dengan <i>Bayesian Information Criterion</i> (BIC) untuk Metode GMM.....	22
9. Perbandingan Citra Uji dengan Citra Hasil Segmentasi EM-GMM.....	25

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis kluster digunakan untuk mengelompokkan objek yang menunjukkan karakteristik serupa dalam kluster yang sama dan objek yang menunjukkan karakteristik berbeda dikelompokkan ke dalam kluster lainnya. Pendekatan *clustering* dibagi menjadi dua kategori yaitu *hard clustering* dan *soft clustering*. Pada *hard clustering*, data dibagi menjadi beberapa kluster sedemikian sehingga setiap elemen data hanya dimiliki oleh satu kluster saja, sedangkan *soft clustering* membentuk kluster sehingga elemen data dapat dimiliki lebih dari satu kluster sekaligus (Grover, 2014).

Salah satu penerapan analisis kluster adalah segmentasi citra digital. Citra atau gambar telah berkembang seiring dengan perkembangan peradaban manusia. Guna menganalisis citra dengan mudah, sangat penting untuk membangun representasi yang lebih sederhana. Proses yang digunakan untuk mempartisi gambar menjadi daerah yang lebih sederhana dan tidak tumpang tindih disebut dengan segmentasi citra (Nguyen & Wu, 2012). Segmentasi citra sangat berguna pada banyak penerapan dalam pengolahan citra dan sistem pengenalan. Hasil dari segmentasi citra berperan sangat penting dalam menyediakan informasi guna diagnosis citra yang lebih baik (Tran *et al.*, 2014).

Terjadi permasalahan saat menerapkan analisis kluster pada segmentasi citra, dimana intensitas piksel akan menarik piksel lainnya yang hampir serupa ke dalam

suatu klaster yang sama. Pada saat melakukan segmentasi citra, metode *hard clustering* akan gagal untuk mencapai konvergen, dikarenakan metode tersebut menempatkan suatu anggota ke dalam satu klaster tertentu saja, apabila suatu data berpindah dari suatu klaster ke dalam klaster lainnya dikarenakan *noise*, maka model klasternya dapat berubah. Permasalahan ini kemungkinan kecil akan jarang terjadi jika menerapkan metode *soft clustering*, karena piksel-piksennya dilengkapi dengan probabilitas atau likelihood dari data, sehingga saat diterapkan ke dalam segmentasi citra menjadi lebih *robust*.

Segmentasi citra akan menjadi lebih baik apabila diterapkan dengan metode *soft clustering*. Salah satu pendekatan *soft clustering* adalah dengan mengikuti bentuk pola distribusi, atau sering dikenal sebagai *model-based clustering*. Klastering berbasis model menggunakan pendekatan berbasis probabilitas yang mengasumsikan bahwa data dihasilkan oleh distribusi probabilitas yang mendasarinya. Gaussian Mixture Model (GMM) adalah pengelompokan berbasis model yang menggunakan distribusi *Gaussian* sebagai komponen dari *mixture model* (Iriawan *et al.*, 2018). Keuntungan dari GMM yaitu memerlukan sejumlah kecil parameter untuk pembelajaran. Keuntungan lainnya adalah parameter ini dapat diestimasi secara efisien dengan mengadopsi algoritma EM untuk memaksimalkan fungsi *log-likelihood* (Nguyen, 2011).

Metode klastering melakukan pemrosesan data untuk menemukan pola dalam data ke dalam beberapa kelompok tanpa dilatih secara eksplisit, hal ini sering disebut sebagai *unsupervised*. Untuk menentukan jumlah klaster yang tepat, diperlukan suatu evaluasi pada model klaster. Evaluasi validasi analisis klaster dapat dilakukan dengan menggunakan nilai *Akaike Information Criteria* (AIC) dan *Bayesian Information Criteria* (BIC). AIC adalah kriteria pemilihan model pertama yang mendapatkan perhatian luas dalam komunitas statistik, dan terus menjadi salah satu alat pemilihan model yang paling banyak dikenal dan digunakan dalam praktik statistik. Menurut Schwarz (1978), BIC terkait erat dengan AIC dan

perbedaan diantaranya adalah BIC memberikan pinalti yang lebih besar untuk jumlah parameter dalam model dibandingkan dengan AIC.

Berdasarkan pemaparan tersebut, penulis tertarik untuk menerapkan metode *Gaussian Mixture Model* (GMM) pada segmentasi citra digital. Algoritma *Expectation-Maximization* (EM) diadopsi sebagai metode optimasi agar parameter GMM dapat diestimasi secara efisien dan mendapatkan hasil estimasi yang lebih baik. Pemilihan jumlah kluster yang optimal akan ditentukan menggunakan nilai *Akaike Information Criteria* (AIC) dan *Bayesian Information Criteria* (BIC).

1.2 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mengestimasi parameter terbaik dari *Gaussian Mixture Model* (GMM) dengan menggunakan *Maximum Likelihood Estimation* (MLE) yang diiterasi dengan algoritma *Expectation-Maximization* (EM).
2. Menerapkan metode *Gaussian Mixture Model* (GMM) pada segmentasi citra digital untuk mendapatkan hasil jumlah kluster yang optimal dengan menggunakan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC).

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah memberikan informasi serta menambah wawasan tentang penerapan ilmu statistika khususnya mengenai penerapan metode *Gaussian Mixture Model* dengan estimasi parameter menggunakan *Maximum Likelihood Estimation* yang diiterasi dengan algoritma *Expectation Maximization* pada segmentasi citra digital.

II. TINJAUAN PUSTAKA

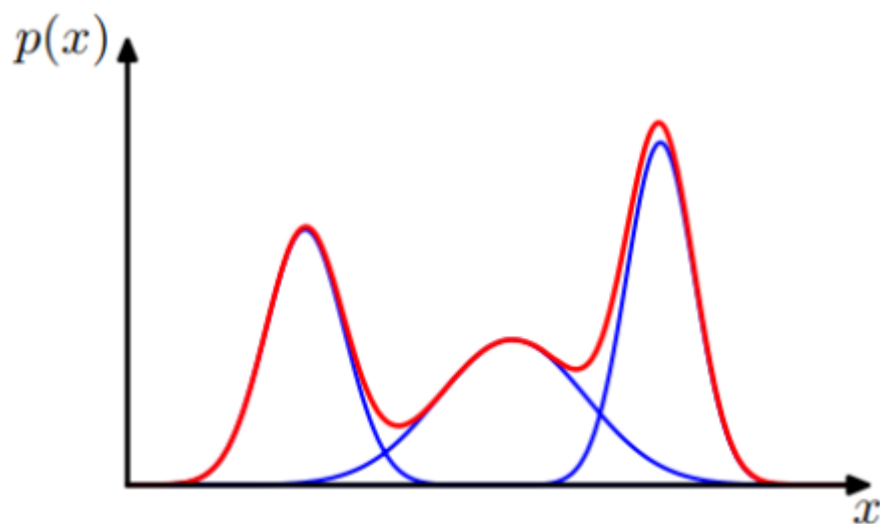
2.1 Analisis Klaster

Analisis klaster digunakan untuk mengelompokkan objek yang menunjukkan karakteristik serupa dalam klaster yang sama dan objek yang menunjukkan karakteristik berbeda dikelompokkan ke dalam klaster lainnya. Pendekatan *clustering* dibagi menjadi dua kategori yaitu *hard clustering* dan *soft clustering*. Pada *hard clustering*, data dibagi menjadi beberapa *cluster* sedemikian sehingga setiap elemen data hanya dimiliki oleh satu *cluster* saja, sedangkan *soft clustering* membentuk *cluster* sehingga elemen data dapat dimiliki lebih dari satu *cluster* berdasarkan tingkat keanggotaannya (Grover, 2014).

Salah satu pendekatan metode *soft clustering* adalah dengan mengikuti bentuk pola distribusi, atau sering dikenal sebagai *model-based clustering*. Metode klastering berbasis model (*model-based clustering*) merupakan suatu pendekatan statistik pada analisis klaster berdasarkan model probabilitas dari suatu data yang berasal dari fungsi distribusi. Dikarenakan data berasal dari campuran dua model probabilitas yang berbeda sehingga hal tersebut menjadi alasan untuk membagi data menjadi dua kelompok atau klaster. Dalam penelitian pengamatan umumnya heterogen, bukan satu kelompok homogen. Model diasumsikan memiliki jumlah klaster yang berhingga, jumlah klaster tetap, dan data berasal dari klaster yang berbeda, sehingga data diasumsikan berasal dari beberapa distribusi probabilitas dalam setiap klaster (Shi, 2005).

2.2 Model Campuran

Distribusi *mixture* banyak digunakan untuk memberikan representasi komputasi yang nyaman untuk memodelkan distribusi data yang kompleks pada fenomena acak (McLachlan *et al.*, 2019). Model campuran (*mixture model*) ditunjukkan dengan jelas pada Gambar 1. Apabila setiap distribusi memiliki parameter sendiri-sendiri, dengan *mean* masing-masing μ_1, μ_2 , dan μ_3 , maka model tersebut dapat dikatakan memiliki tiga kluster dengan centroidnya adalah μ_1, μ_2 , dan μ_3 .



Gambar 1. Fungsi Densitas dari *Mixture Model*

Pada model campuran, setiap komponen distribusi probabilitas sesuai dengan sebuah kluster. Masalah penentuan jumlah kluster dan pemilihan metode *clustering* yang tepat dapat disusun kembali sebagai masalah pemilihan model statistik, dan model yang berbeda dalam jumlah komponen dapat dibandingkan (Fraley & Raftery, 2002). Secara umum, fungsi densitas dari model campuran (*mixture*) didefinisikan dengan persamaan (Iriawan, 2011),

$$f(y|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^K \pi_j f_j(y|\boldsymbol{\theta}_j) \quad (2.1)$$

dimana,

- $f(y|\boldsymbol{\theta}, \boldsymbol{\pi})$: fungsi densitas distribusi *mixture*
- $f_j(y|\boldsymbol{\theta}_j)$: fungsi komponen ke- j
- $\boldsymbol{\theta}_j$: vektor dari setiap parameter yang tidak diketahui dalam distribusi campuran $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K\}$
- $\boldsymbol{\pi}$: vektor parameter proporsi dengan elemen-elemen $\{\pi_1, \pi_2, \dots, \pi_K\}$
- π_j : parameter proporsi komponen distribusi *mixture* dengan $\sum_{j=1}^K \pi_j = 1$ dan $0 \leq \pi_j \leq 1$
- K : jumlah distribusi sebagai bagian dari distribusi campuran yang diberikan

Untuk mengestimasi parameter dari distribusi campuran, banyak pendekatan yang telah dikembangkan, seperti metode momen, metode grafik, metode jarak minimum, metode kemungkinan maksimum, dan metode Bayesian. Tetapi belum tersedia formula eksplisit untuk mengestimasi parameter (McLachlan & Peel, 2000). Metode estimasi kemungkinan maksimum merupakan yang paling umum digunakan untuk mengestimasi distribusi campuran dengan fungsi *likelihood* pada persamaan (2.2) berikut.

$$\begin{aligned} L(y|\boldsymbol{\theta}) &= \prod_{i=1}^N f_{Mixture}(y_i|\boldsymbol{\theta}_j) \\ &= \prod_{i=1}^N \sum_{j=1}^K \pi_j f_j(y_i|\boldsymbol{\theta}_j) \end{aligned} \quad (2.2)$$

Menggunakan fungsi log-likelihood akan menjadi lebih mudah karena konstanta aditif bersifat arbitrer, dengan menggunakan rumus pada persamaan (2.3) berikut.

$$\begin{aligned}
\ell(\boldsymbol{\theta}; y) &= \sum_{i=1}^N \log \left(f_{\text{Mixture}}(y_i | \boldsymbol{\theta}_j) \right) \\
&= \sum_{i=1}^N \log \left(\sum_{j=1}^K \pi_j f_j(y_i | \boldsymbol{\theta}_j) \right)
\end{aligned} \tag{2.3}$$

Secara umum, untuk memaksimalkan fungsi *log-likelihood* ini tidak memiliki solusi bentuk tertutup langsung karena parameternya adalah distribusi campuran dalam logaritma dari model yang mendasarinya. Banyak metode telah diterapkan untuk menyelesaikan masalah fungsi *log-likelihood* yang mengarah ke masalah optimasi non-linier (Shi, 2005).

2.3 Gaussian Mixture Model (GMM)

Gaussian Mixture Model (GMM) adalah pengelompokan berbasis model yang menggunakan distribusi *Gaussian* sebagai komponen dari *mixture model* (Iriawan *et al.*, 2018). Misal diketahui $y_1, y_2, \dots, y_N$ adalah kumpulan piksel pada citra yang diasumsikan memiliki distribusi campuran Gaussian atau biasa disebut dengan *Gaussian Mixture Model* (GMM). Dari persamaan (2.1) dapat diperoleh fungsi densitas GMM sebagai berikut,

$$f(y | \pi, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{j=1}^K \pi_j f(y_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{2.4}$$

dengan fungsi $f(y_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ pada persamaan (2.4) merupakan fungsi kepadatan peluang *Gaussian* multivariat pada persamaan (2.5) berikut,

$$f(y_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_j|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j) \right) \tag{2.5}$$

dimana $\pi = (\pi_1, \dots, \pi_K)$, $0 \leq \pi_j \leq 1$, $j = 1, 2, \dots, K$, $\sum_{j=1}^K \pi_j = 1$. Oleh karena itu, fungsi *joint conditional density* dari kumpulan piksel $y_1, y_2, \dots, y_N$ dapat dimodelkan menjadi,

$$\begin{aligned}
p(y|\pi, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^N f(y_i|\pi, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^N \left[\sum_{j=1}^K \pi_j f(y_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right]
\end{aligned} \tag{2.6}$$

Berdasarkan fungsi *joint conditional density* pada persamaan (2.6), diperoleh fungsi *log-likelihood* dari GMM sebagai berikut,

$$\begin{aligned}
L(\pi, \boldsymbol{\mu}, \boldsymbol{\Sigma}|y) \\
= \sum_{i=1}^N \log \left(\sum_{j=1}^K \frac{\pi_j}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_j|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j) \right) \right)
\end{aligned} \tag{2.7}$$

seperti yang dapat dilihat dari fungsi *likelihood* pada persamaan (2.7), salah satu keunggulan dari metode GMM adalah mempunyai bentuk yang sederhana, dan memerlukan sejumlah kecil parameter (Nguyen, 2011). Permasalahan pada metode GMM yaitu menentukan berapa probabilitas bahwa titik pengamatan akan memasuki klaster tertentu. Besar kecilnya probabilitas tersebut diukur dengan *hidden variable* Z . Nilai ekspektasi dari *hidden variable* tersebut diukur pada tahap E-Step dalam algoritma *Expectation-Maximization* (Dempster *et al.*, 1977).

2.4 Algoritma *Expectation-Maximization* (EM)

Algoritma *Expectation-Maximization* (EM) diperkenalkan pertama kali oleh Dempster, Laird, dan Rubin (1977). Algoritma EM adalah metode iterasi untuk mengestimasi *maximum-likelihood* ketika pengamatan dapat dilihat sebagai data yang tidak lengkap (*incomplete data*). Istilah *incomplete data* dalam bentuk umumnya mengimplikasikan adanya dua ruang sampel dan pemetaan *many-one* dari kedua ruang sampel.

Setiap iterasi dari algoritma EM terdiri dari dua tahap, yaitu *expectation step* (E-step) dan *maximization step* (M-step). Pada E-step nilai ekspektasi dari *likelihood*

dihitung menggunakan parameter data, kemudian M-step menghitung nilai estimasi *maximum-likelihood* dari parameter dengan memaksimalkan nilai ekspektasi dari *likelihood* yang terdapat pada E-step. Kedua tahapan ini dilakukan hingga mencapai nilai konvergen (Dempster *et al.*, 1977).

Algoritma *Expectation-Maximization* (EM) dimulai dari menurunkan fungsi *log-likelihood* pada persamaan (2.7) terhadap μ_j dari komponen *Gaussian*, diperoleh,

$$\begin{aligned} \frac{\partial L(\mu, \Sigma, \pi|y)}{\partial \mu_j} &= \frac{\partial (\sum_{i=1}^N \log(\sum_{j=1}^K \pi_j f_j(y_i | \mu_j, \Sigma_j)))}{\partial \mu_j} \\ 0 &= - \sum_{i=1}^N \frac{\pi_j f_j(y_i | \mu_j, \Sigma_j)}{\sum_{l=1}^K \pi_l f(y_i | \mu_l, \Sigma_l)} \Sigma_j^{-1} (y_i - \mu_j) \end{aligned} \quad (2.8)$$

dimana $\sum_{l=1}^K \pi_l f(y_i | \mu_l, \Sigma_l)$ merupakan variabel laten z_{ij} , dengan probabilitas posterior pada persamaan (2.9) berikut,

$$\begin{aligned} z_{ij}^{(t)} &= f(z_{ij} | y_i, \mu_j^{(t)}, \Sigma_j^{(t)}) \\ &= \frac{f(y_i | z_{ij}, \mu_j^{(t)}, \Sigma_j^{(t)}) f(z_{ij} | \mu_j^{(t)}, \Sigma_j^{(t)})}{f(y_i | \mu_j^{(t)}, \Sigma_j^{(t)})} \\ &= \frac{\pi_j^{(t)} f(y_i | \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{l=1}^K \pi_l^{(t)} f(y_i | \mu_l^{(t)}, \Sigma_l^{(t)})} \end{aligned} \quad (2.9)$$

dimana t menunjukkan langkah iterasi. Penyelesaian dari $\partial L(\mu, \Sigma, \pi|y)/\partial \mu_j = 0$ pada persamaan (2.8) menghasilkan nilai μ_j minimum pada langkah iterasi ke $(t + 1)$.

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^N z_{ij}^{(t)} y_i}{\sum_{i=1}^N z_{ij}^{(t)}} \quad (2.10)$$

Jika fungsi *log-likelihood* pada persamaan (2.7) diturunkan terhadap Σ_j^{-1} dari komponen *Gaussian*, maka diperoleh.

$$[\Sigma_j]^{(t+1)} = \frac{\sum_{i=1}^N z_{ij}^{(t)} (\mathbf{y}_i - \boldsymbol{\mu}_j^{(t+1)}) (\mathbf{y}_i - \boldsymbol{\mu}_j^{(t+1)})^T}{\sum_{i=1}^N z_{ij}^{(t)}} \quad (2.11)$$

Kemudian, dapat diturunkan fungsi *log-likelihood* pada persamaan (2.7) terhadap distribusi *prior* π_j . Pada tahap ini, domain $0 \leq \pi_j \leq 1$ dan $\sum_{j=1}^K \pi_j = 1$ harus dipertimbangkan. Ini dapat dicapai dengan menggunakan *Lagrange multiplier* η dan memaksimalkan turunan berikut,

$$\frac{\partial}{\partial \pi_j} \left[L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi | y) - \eta \left(\sum_{j=1}^K \pi_j - 1 \right) \right] = 0 \quad (2.12)$$

dari persamaan (2.12) selanjutnya didapatkan,

$$\sum_{i=1}^N \frac{f(y_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{l=1}^K \pi_l f(y_i | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)} - \eta = 0 \quad (2.13)$$

Jika kedua sisi pada persamaan (2.13) dikalikan dengan π_j dan digunakan $0 \leq \pi_j \leq 1$ dan $\sum_{j=1}^K \pi_j = 1$, maka diperoleh $\eta = N$. Oleh karena itu, untuk mengeliminasi η dan menyusun ulang turunan terhadap distribusi *prior* π_j diperoleh,

$$\pi_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^N z_{ij}^{(t)} \quad (2.14)$$

(Nguyen, 2011).

Algoritma EM untuk segmentasi dengan GMM dilakukan dengan tahapan berikut:

- a. Inisialisasi parameter *mean* $\boldsymbol{\mu}_j$, *covarians* $\boldsymbol{\Sigma}_j$, dan distribusi *prior* π_j dengan menggunakan metode *k-means*.
- b. Pada E-Step akan dihitung nilai z_{ij} pada persamaan (2.9) dengan nilai parameter yang telah di inisialisasi.
- c. Pada M-Step akan diestimasi ulang parameter $\{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \pi_j\}$.
 - Perbarui nilai *mean* $\boldsymbol{\mu}_j$ dengan menggunakan persamaan (2.10).
 - Perbarui nilai *covarians* $\boldsymbol{\Sigma}_j$ dengan menggunakan persamaan (2.11).
 - Perbarui distribusi *prior* π_j dengan menggunakan persamaan (2.14).

- d. Evaluasi fungsi *log-likelihood* dalam persamaan (2.7) dan periksa konvergensi dari fungsi *log-likelihood* dan nilai parameter. Jika kriteria konvergensi untuk nilai *error* < 0.0001 belum terpenuhi, kembali ke langkah b.

2.5 Pemilihan Model

Pemilihan model merupakan salah satu masalah terpenting dalam analisis kluster campuran berdasarkan distribusi normal campuran multivariat. Pemilihan model meliputi penentuan jumlah komponen (*cluster*) dan pemilihan struktur kovarians yang sesuai dalam analisis cluster campuran. Kriteria informasi sering digunakan dalam pemilihan model dalam analisis kluster campuran. Beberapa kriteria informasi yang umum digunakan dalam literatur adalah AIC dan BIC (Akogul & Erisoglu, 2016).

2.5.1 Akaike's Information Criterion (AIC)

Akaike Information Criterion (AIC) adalah kriteria pemilihan model pertama yang mendapatkan perhatian luas dalam komunitas statistik, dan terus menjadi salah satu alat pemilihan model yang paling banyak dikenal dan digunakan dalam praktik statistik. Pada hal ini, AIC digunakan untuk memperkirakan perkiraan jumlah komponen/kluster yang ada dalam data.

$$AIC = 2p - 2 \ln(L) \quad (2.15)$$

dimana, p adalah jumlah parameter dan L adalah nilai *maximum-likelihood* dari model. Memilih jumlah komponen yang lebih sedikit dalam model adalah langkah yang tepat untuk mengurangi beban komputasi dan *overfitting* (Ragothaman *et al.*, 2016).

2.5.2 *Bayesian Information Criterion (BIC)*

Bayesian Information Criterion (BIC) adalah suatu kriteria untuk pemilihan model. BIC terkait erat dengan *Akaike Information Criterion (AIC)* dan perbedaan diantaranya adalah BIC memberikan pinalti yang lebih besar untuk jumlah parameter dalam model dibandingkan dengan AIC (Schwarz, 1978). Model yang dipilih adalah model yang memiliki nilai BIC paling kecil. BIC untuk GMM dapat dikalkulasi menggunakan persamaan (2.15).

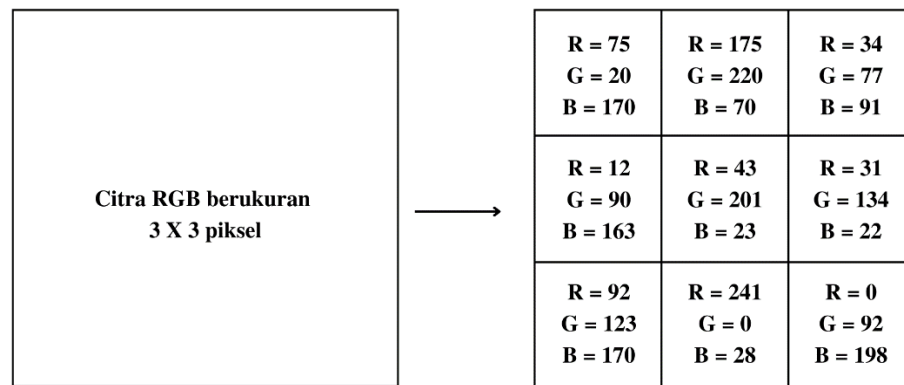
$$BIC = p \ln(N) - 2\ln(L) \quad (2.16)$$

dimana N adalah jumlah observasi.

2.6 Segmentasi Citra Digital

Citra merupakan sinyal multidimensi yang dapat didefinisikan sebagai fungsi dua-dimensi $f(x,y)$, dimana x dan y koordinat spasial, dan amplitudo f pada sembarang pasangan koordinat (x,y) disebut sebagai tingkat keabuan atau *graylevel* (Gonzales & Wood, 2008). Suatu citra mengandung sub-citra yang disebut sebagai *Regions of Interest (ROI)*. Konsep ini menjelaskan fakta bahwa citra mengandung kumpulan objek yang masing-masing dapat menjadi dasar dari suatu *region* atau wilayah (Young *et al.*, 1998).

Menurut Solomon & Breckon (2011), terdapat 3 jenis citra, yaitu citra biner, citra *grayscale*, dan citra RGB. Citra biner merupakan susunan dua dimensi yang menetapkan satu nilai numerik dari himpunan $\{0,1\}$ untuk setiap piksel dalam citra. Citra *grayscale* merupakan susunan dua dimensi yang menetapkan suatu nilai numerik yang umumnya memiliki 256 tingkat keabuan yang berbeda mulai dari hitam hingga putih. Citra RGB merupakan susunan tiga dimensi yang menetapkan tiga nilai numerik untuk setiap piksel, masing-masing nilai sesuai dengan komponen kanal merah, hijau, dan biru (RGB).



Gambar 2. Ilustrasi Citra RGB

Guna menganalisis isi dari citra dengan mudah, sangat penting untuk membangun representasi yang lebih sederhana. Proses yang digunakan untuk mempartisi gambar menjadi daerah yang lebih sederhana dan tidak tumpang tindih disebut dengan segmentasi citra (Nguyen & Wu, 2012). Segmentasi citra dapat dilakukan melalui 3 macam pendekatan. Pendekatan regional bertujuan untuk membagi citra ke dalam wilayah-wilayah sehingga muncul suatu wilayah sesuai kriteria yang diinginkan. Pendekatan batas bertujuan untuk memperoleh batas-batas yang ada antar wilayah. Pendekatan tepi bertujuan untuk mengidentifikasi piksel tepi dan menghubungkan piksel-piksel tersebut ke dalam batas yang diinginkan (Castleman, 1996). Segmentasi citra sangat berguna pada banyak penerapan dalam pengolahan citra dan sistem pengenalan. Hasil dari segmentasi citra berperan penting dalam menyediakan informasi guna diagnosis citra yang lebih baik (Tran *et al.*, 2014).

Histogram citra adalah grafik yang menunjukkan sebaran nilai intensitas piksel pada suatu citra. Dengan membuat histogram citra, informasi penting tentang isi suatu citra digital dapat diketahui. Frekuensi relatif terjadinya intensitas pada citra juga dapat dilihat dari histogram (Munir, 2004). Pada penelitian ini, analisis kluster yang digunakan adalah *model-based clustering* atau pengelompokan berbasis model. Pola *mixture* dari model dapat dilihat dengan membentuk histogram citra, yang selanjutnya dapat dilakukan estimasi awal jumlah *mean* yang digunakan dalam mengestimasi model.

III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2021/2022 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2. Data Penelitian

Data yang digunakan dalam penelitian ini adalah citra buah apel yang diperoleh dari situs www.thedailymeal.com dengan tipe file JPG. Citra uji memiliki dimensi berukuran 870×565 piksel dan berada pada ruang warna RGB yang akan dianalisis menggunakan bahasa pemrograman *Python* versi 3.9.9 dengan *Software Jupyter Notebook* versi 6.1.4. Alat yang digunakan dalam penyelesaian penelitian ini adalah dengan spesifikasi Prosesor Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz ~ 2.71GHz, HDD 1 TB, RAM 4 GB, dan OS Windows 10 Pro 64-bit.

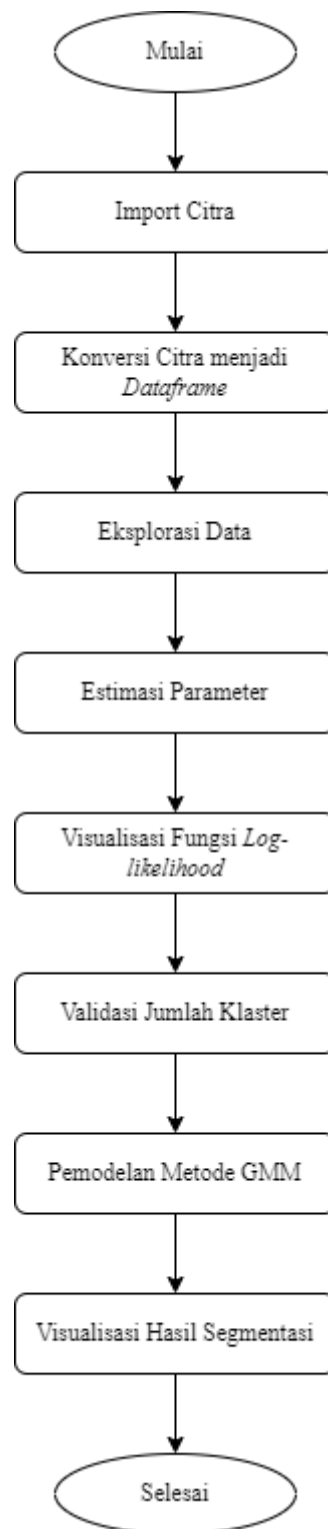
3.3. Metode Penelitian

Kajian pada penelitian ini adalah penerapan metode *Gaussian Mixture Model* (GMM) dengan estimasi parameter menggunakan algoritma *Expectation-*

Maximization (EM) pada segmentasi citra digital. Adapun tahapan penelitian yang akan dilakukan adalah sebagai berikut:

1. Import citra buah apel berukuran 870×565 piksel dengan menggunakan program *Python*.
2. Mengonversikan citra menjadi *dataframe* yang bertujuan untuk menyusun data yang akan digunakan dalam proses analisis kluster dengan metode GMM.
3. Melakukan eksplorasi data dari citra dengan visualisasi data menggunakan grafik densitas kernel untuk melakukan pendugaan awal bahwa data yang digunakan adalah berpola *mixture*.
4. Melakukan estimasi parameter menggunakan *Maximum Likelihood Estimation* (MLE) yang di iterasi dengan algoritma EM untuk GMM dengan tahapan sebagai berikut.
 - a. Inisialisasi parameter *mean* μ_j , *covarians* Σ_j , dan distribusi *prior* π_j dengan menggunakan metode *k-means*.
 - b. Pada E-Step akan dihitung nilai z_{ij} pada persamaan (2.9) dengan nilai parameter yang telah di inisialisasi.
 - c. Pada M-Step akan diestimasi ulang parameter $\{\mu_j, \Sigma_j, \pi_j\}$.
 - Perbarui nilai *mean* μ_j dengan menggunakan persamaan (2.10).
 - Perbarui nilai *covarians* Σ_j dengan menggunakan persamaan (2.11).
 - Perbarui distribusi *prior* π_j dengan menggunakan persamaan (2.14).
 - d. Evaluasi fungsi *log-likelihood* pada persamaan (2.7) dan periksa konvergensi dari fungsi *log-likelihood* dan nilai parameter. Jika kriteria konvergensi untuk nilai *error* < 0.0001 belum terpenuhi, kembali ke langkah b.
5. Membuat visualisasi proses mencari *maximum log-likelihood* di setiap iterasinya.
6. Validasi jumlah kluster terbaik dengan *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC).
7. Membuat pemodelan dengan parameter yang telah diestimasi dengan menggunakan algoritma EM.
8. Membuat visualisasi perbandingan citra uji dengan hasil segmentasi dengan menggunakan EM-GMM.

Diagram alir pada Gambar 3 berikut merupakan langkah-langkah analisis pada penelitian ini.



Gambar 3. Diagram Alir Penelitian

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan uji coba, hasil, dan pembahasan yang telah dilakukan pada bab sebelumnya, dapat ditarik beberapa kesimpulan dari penelitian ini yaitu sebagai berikut:

1. Parameter dari *Gaussian Mixture Model* (GMM) dapat dengan baik diestimasi dengan *Maximum Likelihood Estimation* (MLE) yang diiterasi dengan algoritma *Expectation-Maximization* (EM). Hal ini diindikasikan dengan fungsi *log-likelihood* berhasil mencapai maksimum dan konvergen dengan tingkat toleransi *error* sebesar 0,0001 yang di uji coba pada jumlah kluster 2 hingga 10 kluster. Jumlah kluster juga sangat mempengaruhi banyaknya iterasi untuk mencapai konvergen, semakin banyak jumlah kluster maka semakin banyak pula iterasi yang dibutuhkan.
2. Analisis kluster dengan metode GMM mampu dengan baik mensegmentasi citra sebanyak 7 kluster. Validasi jumlah kluster pada segmentasi citra tersebut berdasarkan pada nilai *Akaike's Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC). Pada penelitian ini, nilai AIC dan BIC terus mengalami penurunan seiring dengan meningkatnya jumlah kluster, sehingga pemilihan jumlah kluster dilihat pada nilai AIC dan BIC yang sedikit mengalami kenaikan pada jumlah kluster 8. Oleh karena itu, pada penelitian ini dipilih jumlah kluster sebelumnya yaitu 7 kluster sebagai yang terbaik bagi metode GMM.

DAFTAR PUSTAKA

- Akogul, S., & Erisoglu, M. 2016. A comparison of information criteria in clustering based on mixture of multivariate normal distributions. *Mathematical and Computational Applications*. **21**(3): 34.
- Castleman, K. R. 1996. *Digital Image Processing*. Prentice Hall, New Jersey.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*. **39**(1): 1-22.
- Fraley, C., & Raftery, A. E. 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*. **97**(458): 611-631.
- Gonzales, R. C., & Woods, R. E. 2008. *Digital image processing*. Pearson Prentice Hall, New Jersey.
- Grover, N. 2014. A Study of Various Fuzzy Clustering Algorithms. *International Journal of Engineering Research*. **3**(3): 177-181.
- Iriawan, N. 2011. Pemodelan Mixture of Mixture dalam Pemilihan Portfolio, hlm. 1-16. Di dalam *Prosiding Seminar Nasional Statistika Universitas Diponegoro 2011*. Program Studi Statistika FMIPA Undip, Surabaya.
- Iriawan, N., Pravitasari, A. A., Fithriasari, K., Irhamah, Purnami, S. W., & Ferriastuti, W. 2018. Comparative Study of Brain Tumor Segmentation using Different Segmentation Techniques in Handling Noise, hlm. 289-293. Di dalam *2018 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, Surabaya.
- Ji, Z., Huang, Y., Sun, Q., & Cao, G. 2016. A Spatially Constrained Generative Asymmetric Gaussian Mixture Model for Image Segmentation. *Journal of Visual Communication and Image Representation*. **40**: 611-626.

- Ji, Z., Huang, Y., Xia, Y., & Zheng, Y. 2017. A Robust Modified Gaussian Mixture Model with Rough Set for Image Segmentation. *Neurocomputing*. **266**: 550-565.
- McLachlan, G. J., & Peel, D. 2000. *Finite Mixture Models, Wiley Series in Probability and Statistics*. John Wiley & Sons Inc., New York.
- McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. 2019. Finite Mixture Models. *Annual Review of Statistics and Its Application*. **6**: 355-378.
- Munir, R. 2004. *Pengolahan Citra Digital Dengan Pendekatan Algoritmik*. Informatika, Bandung.
- Nguyen, T. M. 2011. *Gaussian Mixture Model Based Spatial Information Concept for Image Segmentation*. University of Windsor, Ontario.
- Nguyen, T. M., & Wu, Q. J. 2012. Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*. **23**(4): 621-635.
- Ragothaman, S., Narasimhan, S., Basavaraj, M. G., & Dewar, R. 2016. Unsupervised segmentation of cervical cell images using Gaussian mixture model, hlm. 70-75. Di dalam *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*.
- Schwarz, G. 1878. Estimating the dimension of a model. *The Annals of Statistics*, **6**(2): 461-464.
- Shi, H. S. 2005. Model-based Clustering. Thesis. Statistical Computing University of Waterloo, Ontario.
- Solomon, C., & Breckon, T. 2011. *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, New York.
- Tan, P., Steinbach, M., & Kumar, V. 2006. *Introduction to Data Mining*. Pearson Education, New York.
- Tran, K. A., Vo, N. Q., Nguyen, T. T., & Lee, G. 2014. Gaussian Mixture Model Based on Hidden Markov Random Field for Color Image Segmentation, hlm. 189-197. Di dalam Jeong, Y. S., Park, Y. H., Hsu, C. H., & Park, J. (Penyunting). *Ubiquitous Information Technologies and Applications*. Springer, Berlin.

Wang, L. 2016. Discovering Phase Transitions with Unsupervised Learning. *Physical Review B*. **94**(19): 195105.

Young, I. T., Gerbrands, J. J., & Van Vliet, L. J. 1998. *Fundamentals of image processing*. Delft University of Technology, Delft.