

# Pemodelan Gambar Menggunakan Copula Gaussian dengan Metode Partisi

SRI WINARNI<sup>1</sup>, SAPTO WAHYU INDRATNO<sup>2</sup>, KURNIA NOVITA SARI<sup>3</sup>

<sup>1,2,3</sup>Program Studi Matematika, Fakultas MIPA Institut Teknologi Bandung, Indonesia

<sup>1</sup>Program Studi Statistika, Fakultas MIPA Universitas Padjadjaran, Indonesia

e-mail: sri.winarni@s.itb.ac.id, sri.winarni@unpad.ac.id

## ABSTRAK

Penelitian ini memodelkan gambar menggunakan copula Gaussian. Metode pemodelan yang bersifat fleksibel karena tidak mensyaratkan distribusi normal pada nilai piksel gambar. Masalah kompleksitas komputasi yang disebabkan oleh dimensi data yang besar akan diatasi dengan metode partisi yang dilakukan pada penelitian ini. Data training berupa gambar apel dipartisi menjadi empat bagian yang nantinya akan menjadi variabel bebas dalam model copula Gaussian. Optimasi model dilakukan dengan metode maksimum likelihood dan didapatkan hasil model copula Gaussian dengan *hyperparameter length scale* 1. Metode partisi dapat mereduksi dimensi data sehingga mampu mengatasi permasalahan kompleksitas komputasi.

Kata Kunci: Pemodelan Gambar, Copula Gaussian, Metode Partisi, Reduksi Dimensi Data, Kompleksitas Komputasi.

## ABSTRACT

This study modeled the image using a Gaussian copula. A modeling method that is flexible because it does not require a normal distribution of image pixel values. The problem of computational complexity caused by large data dimensions will be overcome by the partition method used in this study. The training data in the form of an apple image is partitioned into four parts which will later become independent variables in the Gaussian copula model. The optimization of the model was carried out using the maximum likelihood method and the results of the Gaussian copula model with hyperparameter length scale 1. Partition method can reduce the dimensions of the data so that it is able to overcome the problem of computational complexity.

Keywords: Image Modeling, Gaussian Copula, Partition Method, Data Dimension Reduction, Computational Complexity.

## 1. PENDAHULUAN

Perkembangan ilmu dan teknologi membawa dunia data sains kepada era data besar. Ketersediaan data semakin meningkat dengan cepat, hal ini menuntut perkembangan dalam data sains untuk dapat menganalisis data besar yang tersedia. Analisis data besar dilakukan dengan suatu teknik analisis yang disebut dengan data mining. Aplikasi data mining yang sangat terkenal adalah pembelajaran mesin (*machine learning*). Pada bidang pemrosesan gambar, pembelajaran mesin ini salah satunya berupa pemodelan gambar. Kumpulan data besar berupa gambar obyek tertentu digunakan untuk memodelkan obyek gambar tersebut. Sebagai ilustrasi data training gambar apel diberikan pada Gambar 1.



**Gambar 1** Ilustrasi data training gambar apel. Gambar yang digunakan dalam jumlah besar untuk melakukan pembelajaran mesin membentuk model optimum.

Pada pemodelan gambar, data input yang digunakan berupa intensitas ketajaman piksel. Kopoor dkk. (2009) menggunakan model proses Gaussian dalam memodelkan obyek gambar. Intensitas ketajaman gambar pada setiap lokasi piksel disyaratkan berdistribusi Gaussian. Pada prakteknya, intensitas ketajaman gambar ini tidak selalu membentuk distribusi Gaussian, distribusi yang terbentuk bisa saja beta, gamma, weibull, dan lain sebagainya. Pada model proses Gaussian, data yang digunakan adalah pendekatan piksel, artinya data intensitas piksel yang langsung menjadi input analisis. Pada banyak kasus sering terjadi perbedaan yang ekstrim dari nilai intensitas ketajaman gambar yang berpengaruh terhadap analisis. Pendekatan lain yang lebih umum diperlukan untuk dapat memodelkan gambar dengan lebih tepat.

Xuexing (2008) menyebutkan bahwa pendekatan piksel memang lebih mudah dalam analisisnya, tetapi tidak *robust* (tegar) terhadap nilai ekstrim. Pendekatan distribusi menjadi pilihan lain karena bersifat *robust*. Pendekatan ini tidak menggunakan data piksel secara langsung, tetapi menggunakan nilai peluang kumulatif distribusi dari data piksel tersebut. Berdasarkan kendala tersebut maka proses copula Gaussian dapat dijadikan sebagai alternatif analisis.

Beberapa penelitian telah dilakukan untuk memodelkan obyek gambar. Stitou (2009), Salinas dkk. (2011), Lasmar dan Berhoumieu (2014), dan Bansal dkk. (2014) memodelkan gambar menggunakan model copula Gaussian. Permasalahan yang dihadapi dalam pemodelan gambar menggunakan copula Gaussian adalah besarnya dimensi data yang digunakan. Semakin besar ukuran gambar yang digunakan maka dimensi data yang digunakan juga akan semakin besar. Hal ini akan berpengaruh pada kompleksitas komputasi dalam analisis data.

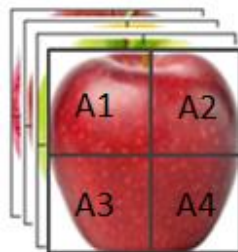
Pada penelitian ini akan digunakan model copula Gaussian dengan metode partisi untuk dapat mereduksi dimensi data sehingga kompleksitas komputasi dapat diatasi. Tujuan dari penelitian ini adalah membentuk model copula Gaussian optimum dengan metode partisi pada kasus data apel. Metode optimasi yang digunakan adalah maksimum likelihood.

## 2. METODE PENELITIAN

Data training yang digunakan pada penelitian ini adalah gambar apel sebanyak 200 yang terdiri dari 100 gambar apel merah dan 100 apel hijau. Analisis dilakukan pada citra abu-abu, sehingga nilai intensitas piksel bernilai 0 – 255. Tahapan analisis yang dilakukan adalah sebagai berikut :

### a. Pra-proses data dan partisi matriks

Sebelum masuk dalam analisis terlebih dahulu dilakukan pre-proses data dengan menormalisasi gambar. Gambar dipotong dengan ukuran yang sama, yaitu  $100 \times 100$ . Selanjutnya ekstraksi gambar dilakukan dengan membaca gambar dan mengekstraksinya kedalam sebuah matriks dengan elemennya adalah nilai intensitas piksel. Satu gambar akan menghasilkan satu matriks intensitas piksel untuk citra abu-abu berukuran  $100 \times 100$ . Matriks tersebut kemudian dibentuk dalam satu vektor baris berukuran  $1 \times 10000$ , sehingga dengan data training sebanyak 200 gambar akan menghasilkan matriks berukuran  $200 \times 10000$ . Partisi dilakukan pada gambar dengan sehingga terbentuk 2 baris dan 2 kolom (Gambar 2).



**Gambar 2** Partisi gambar dalam 2 baris dan 2 kolom

Partisi ini akan memecah matriks menjadi 4 bagian yang masing-masing berukuran  $200 \times 2500$ .

b. Pembentukan model copula Gaussian

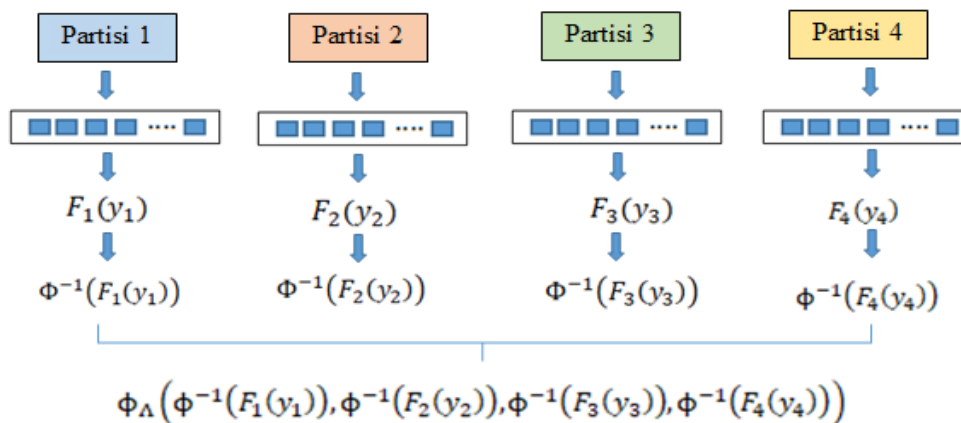
Fungsi distribusi kumulatif copula Gaussian diberikan sebagai berikut :

$$C_{\Lambda}(u_1, u_2, \dots, u_n) = \Phi_{\Lambda}(\Phi^{(-1)}(u_1), \Phi^{(-1)}(u_2), \dots, \Phi^{(-1)}(u_n)) \quad (1)$$

dengan  $\Phi_{\Lambda}$  adalah fungsi distribusi kumulatif multivariat normal dengan rata-rata  $\mathbf{0}$  dan matriks kovariansi  $\Lambda$  yang dibentuk dari fungsi kovariansi (kernel)  $\Lambda_{ij} = k(x_i, x_j)$ . Selanjutnya  $\Phi$  merupakan fungsi distribusi kumulatif normal baku dengan  $u_i = F_i(y_i)$ ,  $i = 1, 2, \dots, n$ ,  $F_i(y_i)$  merupakan distribusi marginal kumulatif, dan  $\Phi^{(-1)}$  merupakan invers normal baku (Nelsen, 2006). Fungsi kovariansi yang digunakan pada penelitian ini adalah eksponensial kuadrat yang diberikan pada Persamaan 2.

$$k(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2} |x_p - x_q|^2\right) \quad (2)$$

dengan  $\mathbf{x}$  adalah input lokasi piksel,  $\sigma_f^2$  signal variance dan  $\ell$  length scale yang merupakan hyperparameter (Rasmussen, 2016). Pembentukan model copula Gaussian dilakukan dengan skema Gambar 3.



**Gambar 3** Skema pembentukan model copula Gaussian

Nilai piksel pada satu partisi akan menjadi satu vektor yang merupakan variabel bebas pada model copula Gaussian. Penentuan distribusi marginal dilakukan dengan uji Anderson Darling dengan hipotesis sebagai berikut :

$H_0$  : data mengikuti distribusi tertentu

$H_1$  : data tidak mengikuti distribusi tertentu

Statistik uji yang digunakan adalah :

$$W_n^2 = -n - \frac{1}{n} \sum_{j=1}^n (2j - 1) [\ln(u_j) + \ln(1 - u_{n-j+1})] \quad (3)$$

Jika data tidak mengikuti distribusi teoritik tertentu maka dapat digunakan distribusi empirik.

c. Optimasi hyperparameter

Optimasi dilakukan dengan maksimum log likelihood yang diberikan pada Persamaan 4.

$$L(\theta) = -\frac{1}{2} \log(|\Lambda_n|) + \frac{1}{2} z_n^T (I - \Lambda_n^{-1}) z_n + \sum_{i=1}^n \log(f_i(y_i)) \quad (4)$$

dengan  $Z_n$  merupakan invers normal baku.

### 3. HASIL DAN PEMBAHASAN

#### Hasil partisi matriks

Hasil ekstraksi dari sebuah gambar berukuran  $p \times p$  adalah matriks intensitas piksel berukuran  $p \times p$  yang diberikan sebagai berikut :

$$M_{p \times p} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix}$$

dengan  $a_{ij}$  adalah intensitas ketajaman gambar pada lokasi piksel baris ke- $i$  kolom ke- $j$ . Matriks ini kemudian diubah dalam bentuk vektor berukuran  $1 \times p^2$  yang diberikan sebagai berikut :

$$A_{1 \times p^2} = [a_{11} \quad a_{12} \quad \dots \quad a_{1p} \quad a_{21} \quad a_{22} \quad \dots \quad a_{2p} \quad \dots \quad a_{pp}]$$

Sehingga dengan data training sebanyak 200 akan menghasilkan matriks intensitas piksel sebagai berikut :

$$X_{n \times p^2} = \begin{bmatrix} a_{111} & \dots & a_{11p} & a_{121} & \dots & a_{12p} & \dots & \dots & a_{1p1} & \dots & a_{1pp} \\ a_{211} & \dots & a_{21p} & a_{221} & \dots & a_{22p} & \dots & \dots & a_{2p1} & \dots & a_{2pp} \\ a_{311} & \dots & a_{31p} & a_{321} & \dots & a_{32p} & \dots & \dots & a_{3p1} & \dots & a_{3pp} \\ a_{411} & \dots & a_{41p} & a_{421} & \dots & a_{42p} & \dots & \dots & a_{4p1} & \dots & a_{4pp} \\ a_{511} & \dots & a_{51p} & a_{521} & \dots & a_{52p} & \dots & \dots & a_{5p1} & \dots & a_{5pp} \\ a_{611} & \dots & a_{61p} & a_{621} & \dots & a_{62p} & \dots & \dots & a_{6p1} & \dots & a_{6pp} \\ a_{711} & \dots & a_{71p} & a_{721} & \dots & a_{72p} & \dots & \dots & a_{7p1} & \dots & a_{7pp} \\ a_{811} & \dots & a_{81p} & a_{821} & \dots & a_{82p} & \dots & \dots & a_{8p1} & \dots & a_{8pp} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ a_{n11} & \dots & a_{n1p} & a_{n21} & \dots & a_{n2p} & \dots & \dots & a_{np1} & \dots & a_{npp} \end{bmatrix}$$

Elemen  $a_{kij}$  adalah nilai ketajaman gambar pada gambar ke- $k$ , lokasi piksel baris ke- $i$ , kolom ke- $j$ , dengan  $k = 1, 2, \dots, n$ ;  $i = 1, 2, \dots, p$  dan  $j = 1, 2, \dots, p$ . Notasi  $n$  menunjukkan banyak gambar dan  $p$  adalah ukuran baris dan kolom piksel. Pada penelitian ini  $n = 200$  dan  $p = 100$ . Matriks  $X_{n \times p^2}$  ini akan dipartisi sesuai partisi gambar dan didapatkan partisi matriks sebagai berikut :

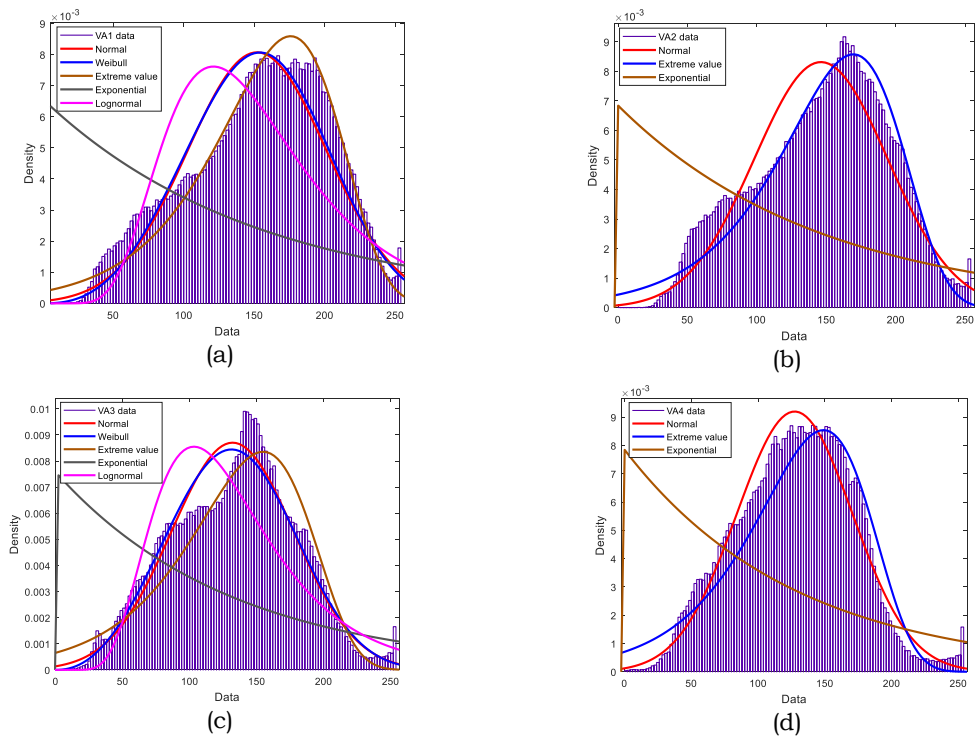
$$X_{n \times p^2} = \begin{bmatrix} a_{1(1)1} & \dots & a_{1(1)\frac{p}{2}} & a_{1(1)\frac{p}{2}+1} & \dots & a_{1(1)p} & \dots & \dots & a_{1(\frac{p}{2})1} & a_{1(\frac{p}{2})1+1} & \dots & a_{1(\frac{p}{2})\frac{p}{2}} & a_{1(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{1(\frac{p}{2})p} & \dots & \dots & a_{1(p)p} \\ a_{2(1)1} & \dots & a_{2(1)\frac{p}{2}} & a_{2(1)\frac{p}{2}+1} & \dots & a_{2(1)p} & \dots & \dots & a_{2(\frac{p}{2})1} & a_{2(\frac{p}{2})1+1} & \dots & a_{2(\frac{p}{2})\frac{p}{2}} & a_{2(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{2(\frac{p}{2})p} & \dots & \dots & a_{2(p)p} \\ a_{3(1)1} & \dots & a_{3(1)\frac{p}{2}} & a_{3(1)\frac{p}{2}+1} & \dots & a_{3(1)p} & \dots & \dots & a_{3(\frac{p}{2})1} & a_{3(\frac{p}{2})1+1} & \dots & a_{3(\frac{p}{2})\frac{p}{2}} & a_{3(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{3(\frac{p}{2})p} & \dots & \dots & a_{3(p)p} \\ a_{4(1)1} & \dots & a_{4(1)\frac{p}{2}} & a_{4(1)\frac{p}{2}+1} & \dots & a_{4(1)p} & \dots & \dots & a_{4(\frac{p}{2})1} & a_{4(\frac{p}{2})1+1} & \dots & a_{4(\frac{p}{2})\frac{p}{2}} & a_{4(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{4(\frac{p}{2})p} & \dots & \dots & a_{4(p)p} \\ a_{5(1)1} & \dots & a_{5(1)\frac{p}{2}} & a_{5(1)\frac{p}{2}+1} & \dots & a_{5(1)p} & \dots & \dots & a_{5(\frac{p}{2})1} & a_{5(\frac{p}{2})1+1} & \dots & a_{5(\frac{p}{2})\frac{p}{2}} & a_{5(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{5(\frac{p}{2})p} & \dots & \dots & a_{5(p)p} \\ a_{6(1)1} & \dots & a_{6(1)\frac{p}{2}} & a_{6(1)\frac{p}{2}+1} & \dots & a_{6(1)p} & \dots & \dots & a_{6(\frac{p}{2})1} & a_{6(\frac{p}{2})1+1} & \dots & a_{6(\frac{p}{2})\frac{p}{2}} & a_{6(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{6(\frac{p}{2})p} & \dots & \dots & a_{6(p)p} \\ a_{7(1)1} & \dots & a_{7(1)\frac{p}{2}} & a_{7(1)\frac{p}{2}+1} & \dots & a_{7(1)p} & \dots & \dots & a_{7(\frac{p}{2})1} & a_{7(\frac{p}{2})1+1} & \dots & a_{7(\frac{p}{2})\frac{p}{2}} & a_{7(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{7(\frac{p}{2})p} & \dots & \dots & a_{7(p)p} \\ a_{8(1)1} & \dots & a_{8(1)\frac{p}{2}} & a_{8(1)\frac{p}{2}+1} & \dots & a_{8(1)p} & \dots & \dots & a_{8(\frac{p}{2})1} & a_{8(\frac{p}{2})1+1} & \dots & a_{8(\frac{p}{2})\frac{p}{2}} & a_{8(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{8(\frac{p}{2})p} & \dots & \dots & a_{8(p)p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ a_{n(1)1} & \dots & a_{n(1)\frac{p}{2}} & a_{n(1)\frac{p}{2}+1} & \dots & a_{n(1)p} & \dots & \dots & a_{n(\frac{p}{2})1} & a_{n(\frac{p}{2})1+1} & \dots & a_{n(\frac{p}{2})\frac{p}{2}} & a_{n(\frac{p}{2})\frac{p}{2}+1} & \dots & a_{n(\frac{p}{2})p} & \dots & \dots & a_{n(p)p} \end{bmatrix}$$

Partisi matriks ini yang akan membentuk model copula Gaussian.

**Hasil pembentukan model copula Gaussian**

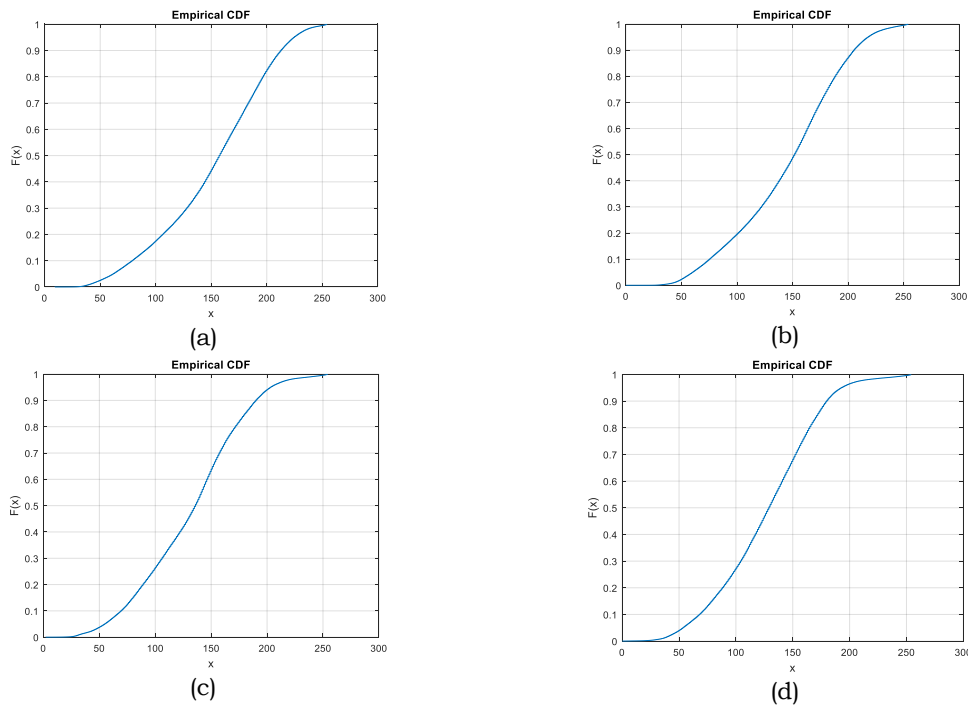
Tahap berikutnya adalah menentukan fungsi distribusi kumulatif dari variabel acak. Fungsi distribusi kumulatif ditentukan dengan uji kecocokan distribusi Anderson Darling menggunakan Persamaan 3. Nilai peluang kumulatif yang didapat dari distribusi-distribusi yang berbeda tersebut ditransformasi menjadi  $\Phi^{-1}[F(y)]$ , nilai inilah yang menjadi input untuk proses copula Gaussian.

Pencocokan distribusi pada masing-masing partisi dilakukan menggunakan distribusi normal, weibull, extreme value, exponential dan lognormal. Hasil pencocokan distribusi diberikan pada Gambar 4.



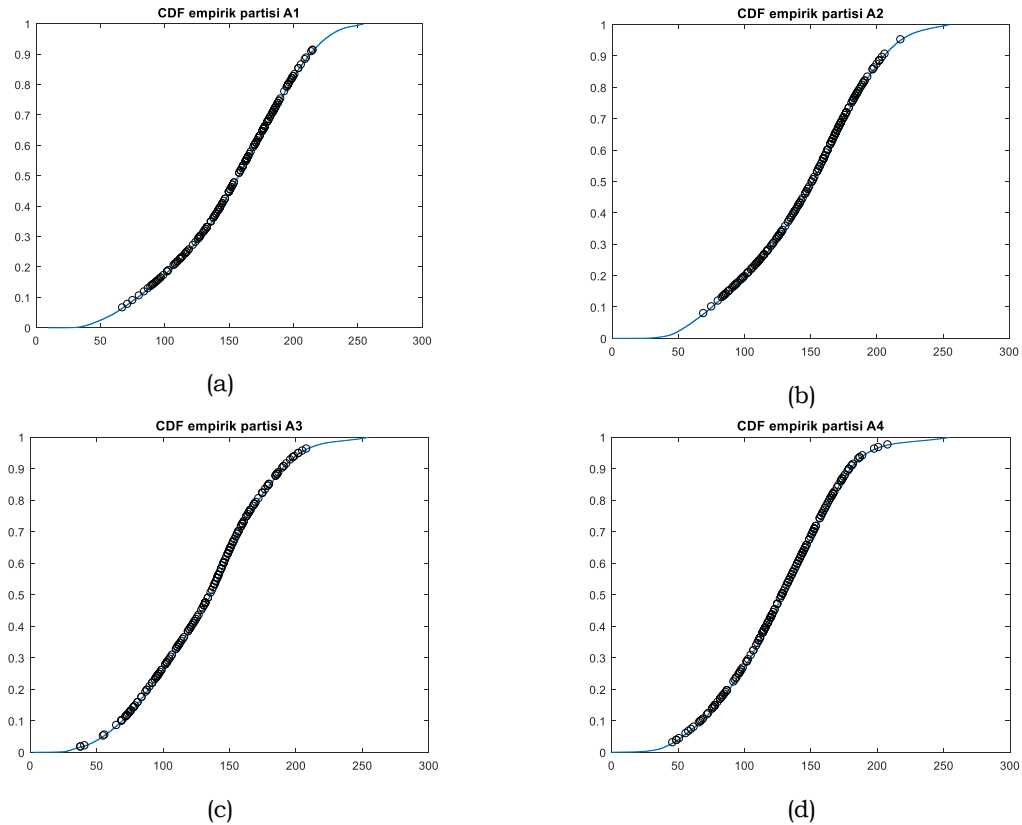
**Gambar 4** Pencocokan distribusi data partisi 1 (a), partisi 2 (b), partisi 3 (c) dan partisi 4 (d)

Hasil pencocokan distribusi pada Gambar 4 menunjukkan bahwa distribusi yang digunakan tidak cocok dengan data. Pengujian dilakukan dengan uji Anderson Darling menggunakan Persamaan 3 dan didapatkan nilai p-value  $> 0.05$  sehingga disimpulkan bahwa data tidak mengikuti distribusi yang diujikan. Berhubung distribusi yang diujikan tidak cocok dengan data, maka pendekatan yang digunakan adalah distribusi empirik. Hasil distribusi empirik kumulatif dari data partisi skema A diberikan pada Gambar 5.



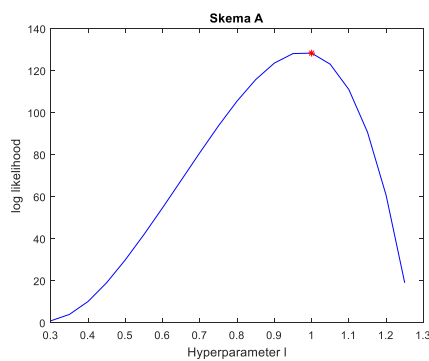
**Gambar 5** Distribusi empirik kumulatif partisi 1 (a), partisi 2 (b), partisi 3 (c), dan partisi 4 (d).

Pola distribusi empirik kumulatif yang terbentuk pada Gambar 5 berbeda untuk setiap partisi. Selanjutnya nilai distribusi empirik kumulatif pada nilai rata-rata piksel setiap partisi gambar diberikan pada Gambar 6.



**Gambar 6** Nilai distribusi empirik kumulatif pada rata-rata piksel partisi 1 (a), partisi 2 (b), partisi 3 (c), dan partisi 4 (d).

Berdasarkan hasil yang didapat pada Gambar 6 akan dilakukan pendugaan parameter model proses copula Gaussian menggunakan Persamaan 3, didapatkan hasil sebagai berikut :



**Gambar 7** Pendugaan hyperparameter model proses copula Gaussian

Hasil pendugaan *hyperparameter* pada Gambar 7 bahwa *hyperparameter* optimum didapatkan pada nilai 1. Artinya bahwa model copula Gaussian optimum untuk gambar apel adalah model copula Gaussian dengan *hyperparameter length scale* sebesar 1.

#### 4. SIMPULAN DAN SARAN

Pada pemodelan gambar apel, penggunaan empat variabel bebas dari metode partisi dapat mereduksi dimensi data dan mendapatkan model optimum dengan hyperparameter sebesar 1. Penelitian lebih lanjut dapat dilakukan untuk klasifikasi dan evaluasi menggunakan model tersebut. Mendeteksi gambar baru dan mengklasifikasikannya sebagai gambar apel atau bukan

#### UCAPAN TERIMA KASIH.

Ucapan terima kasih kepada Program Penelitian, Pengabdian kepada Masyarakat dan Inovasi ITB (*P3MI-ITB*) 2021 yang telah mendanai penelitian ini.

#### DAFTAR PUSTAKA

- Bansal, R., Hao, X., Liu, J. dan Peterson, B.S. (2014) : Using Copula Distributions to Support More Accurate Imaging-based Diagnostic Classifiers for Neuropsychiatric Disorders. *Magnetic Resonance Imaging*, 32. 1102-1113.
- Kapoor, A., Grauman, K., Urtasun, R. dan Darrel, T. (2009) : Gaussian Process for Object Categorization, *Int J Comput Vis*, 10.1007.
- Lasmar, N.E. dan Berthoumieu, Y.A. (2014) : Gaussian Copula Multivariate Modeling for Image Texture Retrieval Using Wavelet Transforms, *Journal & Magazine IEEE Transactions on Image Processing*, 23(5), 2246-2261.
- Nelsen, R.B. (2006) : *An Introduction to Copulas*, Springer, Second edition.
- Rasmussen, C.E. dan Williams, C.K.I. (2006) : *Gaussian Processes for Machine Learning*, The MIT Press Cambridge.
- Salinas, R., Hernandez, A., Rivera, M.J.J. dan Villa, E.R. (2011) : Using Gaussian Copulas in Supervised Probabilistic Classification, *Springer-Verlag Berlin Heidelberg*. 328. 355-372.
- Stitou, Y., Lasmar, N. dan Berthoumieu, Y. (2009) : Copulas Based Multivariate Gamma Modeling for Texture Classification. IEEE. ICASSP 2009.
- Zuexing, Z. (2008) : Image Change Detection using Copulas, *ICSP Proceedings*, 909-913.