

ANALISIS KLASTERISASI MULTI DISTRIBUSI KONTINU

SKRIPSI



**OLEH
HANDIKO
NIM 2019051014009**

**PROGRAM STUDI MATEMATIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS CENDERAWASIH
JAYAPURA
2023**

ANALISIS KLASTERISASI MULTI DISTRIBUSI KONTINU

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains
Program Studi Matematika**



**OLEH
HANDIKO
NIM 2019051014009**

**PROGRAM STUDI MATEMATIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS CENDERAWASIH
JAYAPURA
2023**

ABSTRAK

Handiko. 2023. **Analisis Klasterisasi Multi Distribusi Kontinu**. Skripsi Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Cenderawasih.

Penelitian ini mengkaji tentang pengklasteran distribusi kontinu dengan dua parameter yaitu parameter bentuk dan parameter skala. Distribusi yang digunakan dalam penelitian ini adalah distribusi Erlang, Fatigue Life, Frechet, Gamma, Log-Logistik, Pareto, Pearson Tipe V dan Weibull. Untuk menentukan distribusi apa saja yang mirip, terlebih dahulu harus dilakukan estimasi parameter. Kemudian nilai parameter yang diperoleh akan digunakan untuk melakukan uji hipotesis. Nilai statistik uji yang diperoleh melalui uji hipotesis, kemudian akan dijadikan dasar dalam melakukan klasterisasi terhadap distribusi kontinu dengan dua parameter. Tujuan dari penelitian ini adalah untuk mengetahui distribusi apa saja yang mirip. Hasil penelitian menunjukkan bahwa untuk data dengan $range \leq 20$ dengan jumlah data ≤ 100 , dan $range \leq 50$ dengan jumlah data ≤ 50 pada model sebaran tertentu, distribusi Gamma, Log-Logistik, dan Weibull memiliki kemiripan. Jumlah klaster optimal untuk data dengan $range$ (1 – 10) sebanyak 50 dan 100 data (data set 1 dan 2), $range$ (1 – 10) sebanyak 300 data (data set 1), $range$ (1 – 20) sebanyak 50 data (data set 2) dan 100 data (data set 1), dan $range$ (1 – 50) sebanyak 50 data (data set 2) yaitu sebanyak 2 klaster. Data dengan $range$ (1 – 10) sebanyak 300 data (data set 2) dan $range$ (1 – 50) sebanyak 50 data (data set 1) yaitu sebanyak 3 klaster. Dan data dengan $range$ (1 – 20) sebanyak 50 data (data set 1) dan 100 data (data set 2) sebanyak 5 klaster.

Kata Kunci: klasterisasi, distribusi kontinu, uji kolmogorov-smirnov, uji anderson-darling

ABSTRACT

Handiko. 2023. **Continuous Multi Distribution Clustering Analysis**. Thesis Department of Mathematics, Faculty of Mathematics and Natural Sciences, Cenderawasih University.

This study examines the continuous distribution clustering with two parameters, namely the shape parameter and the scale parameter. The distributions used in this study are the Erlang, Frechet, Fatigue Life, Gamma, Log-Logistic, Pareto, Pearson Type V, and Weibull distributions. To determine which distributions are similar, parameter estimation must first be carried out to obtain the parameter values. The parameter values obtained will then be used to test the hypothesis. The value of the test statistic obtained through hypothesis testing will then be used as the basis for clustering some continuous distribution with two parameters. The purpose of this study is to find out which distributions are similar. The results of this research show that for data with range ≤ 20 and total data ≤ 100 data, and data with range ≤ 50 and total data ≤ 50 in certain distributions models, the gamma, Log-Logistic, and Weibull distributions are similar. The optimal number of clusters for data with range (1 – 10) with a total of 50 and 100 data (data set 1 and 2), range (1 – 10) with a total of 300 data (data set 1), range (1 – 20) with a total of 50 data (data set 2) and 100 data (data set 1), and range (1 – 50) with a total of 50 data (data set 2) as many as 2 clusters. Data with range (1 – 10) with a total of 300 data (data set 2) and range (1 – 50) with a total of 50 data (data set 1) as many as 3 clusters. and data with range (1 – 20) with a total of 50 data (data set 1) and 100 data (data set 2) as many as 5 clusters.

Keywords: *clustering, continuous distribution, Kolmogorov-smirnov test, Anderson-darling test*

LEMBAR PERSETUJUAN

Skripsi dengan judul: **ANALISIS KLASTERISASI MULTI DISTRIBUSI KONTINU** oleh Handiko telah diperiksa dan disetujui untuk diuji.

Jayapura, 12 Juni 2023

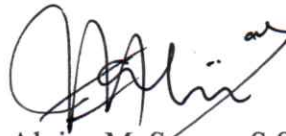
Pembimbing I



Felix Reba, S.Si., M.Sc.

NIP 19850618 201903 1 009

Pembimbing II



Alvian M. Sroyer, S.Si., M.Si.

NIP 19810829 200501 1 001

LEMBAR PENGESAHAN

Skripsi dengan judul: **ANALISIS KLASTERISASI MULTI DISTRIBUSI KONTINU** oleh Handiko telah dipertahankan didepan Dewan Penguji pada tanggal 27 Juni 2023.

Dewan Penguji:

Nama

Jabatan

Tanda Tangan

1. Feby Seru, S.Si., M.Si.

Ketua

(.....)

NIP 19870831 201903 2 015

2. Tiku Tandianga, S.Si., M.Sc.

Anggota

(.....)

NIP 19810415 200604 2 003

3. Abraham, M.Si.

Anggota

(.....)

NIP 19780712 200801 1 006

4. Felix Reba, S.Si., M.Sc.

Anggota

(.....)

NIP 19850618 201903 1 009

5. Alvian M. Sroyer, S.Si., M.Si.

Anggota

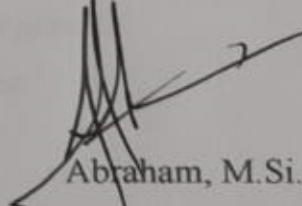
(.....)

NIP 19810829 200501 1 001

Mengetahui,

Ketua Jurusan

Ketua Program Studi



Dr. Jonner Nainggolan, M.Si.

Abraham, M.Si.

NIP.19660702 199303 1 001

NIP.19780712 200801 1 006

Mengesahkan,

Dekan Fakultas MIPA



Dr. Dink Y. P. Runtuboi, M.Kes.

NIP.19760123 200112 1 003

LEMBAR MOTTO DAN PERSEMBAHAN

“Dreams without goals, are just dreams and ultimately they fuel disappointment. On the road to achieving your dreams, you must apply discipline. But more importantly, consistency.”

(Denzel Washington)

“No human is limited”

(Eliud Kipchoge)

“The best time to plant a tree was 20 years ago. The second best time is now.”

(Chinese Proverb)

Skripsi ini kupersembahkan kepada :

- 1. Kedua orang tua penulis, ayah dan ibu tercinta,
Darwan Halim dan Wong Min Gik.*
- 2. Adik penulis yang terkasih.*
- 3. Seluruh sahabat penulis*
- 4. Seluruh pembaca*

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi S1 FMIPA Uncen yang tidak dipublikasikan, terdaftar dan tersedia di Perpustakaan Universitas Cenderawasih dan terbuka untuk umum dengan ketentuan bahwa hak cipta ada pada penulis.

Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau peringkasan hanya dapat dilakukan seizin penulis, dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Memperbanyak atau menerbitkan sebagian atau seluruh skripsi haruslah seizin Rektor Universitas Cenderawasih.

Perpustakaan yang meminjam skripsi ini untuk keperluan anggotanya harus mengisi nama dan tanda tangan peminjam dan tanggal pinjam.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, atas segala berkat dan karunia yang telah diberikan sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Klasterisasi Multi Distribusi Kontinu”.

Penulis menyadari bahwa skripsi ini dapat terselesaikan berkat bantuan baik moral maupun materil dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Dr. Dirk Y. P. Runtuboi, M.Kes., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Cenderawasih, atas kesempatan yang diberikan untuk menjalani studi di Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Dr. Jonner Nainggolan, M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Cenderawasih, atas kesempatan yang diberikan untuk menjalani studi di Jurusan Matematika.
3. Abraham, M.Si., selaku Ketua Program Studi Matematika sekaligus dosen penguji III atas segala waktu, kesabaran, nasehat, dan motivasi yang telah diberikan, sehingga penulisan skripsi ini dapat terselesaikan.
4. Felix Reba, S.Si., M.Sc., selaku dosen wali sekaligus dosen pembimbing I atas segala waktu, kesabaran, nasehat, motivasi dan bimbingannya, sehingga penulisan skripsi ini dapat terselesaikan.
5. Alvian M. Sroyer, S.Si., M.Si., selaku dosen pembimbing II atas segala waktu, kesabaran, nasehat, motivasi dan bimbingannya, sehingga penulisan skripsi ini dapat terselesaikan.
6. Feby Seru, S.Si., M.Si., selaku dosen penguji I atas segala waktu, kesabaran, nasehat, dan motivasi yang telah diberikan, sehingga penulisan skripsi ini dapat terselesaikan.
7. Tiku Tandianga, S.Si., M.Sc., selaku dosen penguji II atas segala waktu, kesabaran, nasehat, dan motivasi yang telah diberikan, sehingga penulisan skripsi ini dapat terselesaikan.
8. Seluruh dosen dan staf FMIPA Uncen, khususnya dosen Jurusan Matematika.
9. Kedua orang tua tercinta, Bapak Darwan Halim dan Ibu Wong Min Gik yang senantiasa mendukung dalam doa dan memberi nasehat serta kasih sayang

yang tulus kepada penulis, sehingga penulis dapat menyelesaikan studi dengan baik.

10. Adik penulis, Willy yang senantiasa mendoakan dan memberikan motivasi kepada penulis, sehingga penulis dapat menyelesaikan studi dengan baik.
11. Sahabat-sahabat penulis, Miftachul Jannah, Sitti Ham Nini, Dal Jahra, Zahra I. P. Astuti, Romauli Simanjuntak, Trimalena Siahaan, Natalia Kawer, Femiance Lemauk, Marice Urbasa yang selalu memberikan semangat dukungan serta doa kepada penulis.
12. Semua pihak yang tidak dapat disebutkan satu persatu yang telah memberikan bantuan baik secara moril maupun materil kepada penulis.

Semoga limpahan rahmat dan kasih sayang Tuhan Yang Maha Esa senantiasa menyertai Bapak, Ibu dan Saudara/i sekalian.

Jayapura, 23 Juni 2023

Penulis

DAFTAR ISI

PEDOMAN PENGGUNAAN SKRIPSI	ii
UCAPAN TERIMA KASIH.....	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR SINGKATAN DAN LAMBANG	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metode Penelitian	4
1.7 Sistematika Penulisan	5
BAB II LANDASAN TEORI.....	6
2.1 Himpunan.....	6
2.2 Fungsi.....	6
2.3 Fungsi Dua Variabel	8
2.4 Limit.....	9
2.5 Turunan.....	10
2.6 Turunan Parsial	11
2.7 Notasi Sigma.....	12
2.8 Integral	12
2.9 Peluang dan Ruang Sampel	14
2.10 Peubah Acak	15
2.11 Populasi dan Sampel	15
2.12 Distribusi Probabilitas.....	15
2.13 Beberapa Distribusi Kontinu.....	17
2.14 Parameter	22

2.15	Metode Newton-Raphson	23
2.16	Metode Estimasi Kemungkinan Maksimum.....	27
2.17	Supremum dan Infimum	30
2.18	Uji Kolmogorov-Smirnov	31
2.19	Uji Anderson-Darling	36
2.20	Normalisasi <i>Z-Score</i>	39
2.21	Analisis Kluster	40
2.22	Single Linkage Clustering.....	40
2.23	Index Davies-Bouldin	47
BAB III	METODOLOGI PENELITIAN	54
3.1	Jenis dan Sumber Data.....	54
3.2	Metode Penelitian	54
3.3	Tahapan Penelitian.....	55
BAB IV	HASIL DAN PEMBAHASAN	57
4.1	Membangkitkan Data.....	57
4.2	Estimasi Parameter dengan Metode Estimasi Kemungkinan Maksimum	57
4.3	Uji Kolmogorov-Smirnov	75
4.4	Uji Anderson-Darling	79
4.5	Membentuk Tabel Statistik Uji.....	84
4.6	Menentukan Kluster yang Terbentuk.....	86
4.7	Interpretasi Data.....	101
BAB V	PENUTUP	103
5.1	Kesimpulan	103
5.2	Saran	108
DAFTAR PUSTAKA		109
LAMPIRAN		111

DAFTAR GAMBAR

Gambar 2.1 Fungsi Dua Variabel	8
Gambar 2.2 Grafik Fungsi Kepadatan Peluang Distribusi Gamma.....	16
Gambar 2.3 Grafik Fungsi Distribusi Kumulatif Distribusi Gamma	17
Gambar 2.4 Grafik Fungsi Kepadatan Peluang Distribusi Normal	23
Gambar 2.5 Penurunan Newton-Raphson	24
Gambar 2.6 Estimasi Parameter dengan Matlab	30
Gambar 2.7 Uji Kolmogorov-Smirnov dengan Matlab.....	35
Gambar 2.8 Uji Anderson-Darling dengan Matlab	39
Gambar 2.9 Evaluasi Kluster dengan R Studio	52
Gambar 3.1 Tahapan Penelitian.....	55
Gambar 4.1 Histogram Data Set 1 dengan Range Data 10 Sebanyak 50 Data .	102
Gambar 4.2 Histogram Data Set 2 dengan Range Data 10 Sebanyak 50 Data .	102
Gambar 5.1 Distribusi Gamma, Log-Logistik, dan Weibull dengan Range 20 Sebanyak 50 Data (Data Set 1)	107
Gambar 5.2 Distribusi Gamma, Log-Logistik, Weibull dengan Range Data 20 Sebanyak 50 Data (Data Set 2)	107

DAFTAR TABEL

Tabel 2.1 Data Acak.....	29
Tabel 2.2 Uji Kolmogorov-Smirnov Distribusi Poisson	32
Tabel 2.3 Uji Anderson Darling	37
Tabel 2.4 Data	41
Tabel 2.5 Jarak Antar Klaster.....	43
Tabel 2.6 Jarak Antar Klaster pada Iterasi Ke-1	44
Tabel 2.7 Jarak Antar Klaster pada Iterasi Ke-2	45
Tabel 2.8 Jarak Antar Klaster pada Iterasi Ke-3	45
Tabel 2.9 Jarak Antar Klaster pada Iterasi Ke-4	46
Tabel 2.10 Hasil Klasterisasi dengan Matlab	47
Tabel 2.11 Klaster yang Terbentuk	49
Tabel 2.12 Titik Pusat.....	50
Tabel 2.13 Ukuran Kemiripan.....	51
Tabel 2.14 Nilai Index Davies-Bouldin.....	53
Tabel 4.1 Hasil Iterasi Newton-Raphson Distribusi Weibull.....	66
Tabel 4.2 Iterasi Newton-Raphson Distribusi Frechet	69
Tabel 4.3 Hasil Estimasi Parameter.....	73
Tabel 4.4 Uji Kolmogorov-Smirnov Distribusi Weibull.....	75
Tabel 4.5 Hasil Uji Kolmogorov-Smirnov.....	78
Tabel 4.6 Uji Anderson-Darling Distribusi Weibull	80
Tabel 4.7 Hasil Uji Anderson-Darling	82
Tabel 4.8 Statistik Uji.....	84
Tabel 4.9 Nilai Statistik Uji Untuk Data Set 1 dengan Range data (1-50) sebanyak 50 data	86
Tabel 4.10 Hasil Standarisasi Data.....	88
Tabel 4.11 Jarak Antar Data.....	89
Tabel 4.12 Jarak Antar Data Iterasi-1.....	90
Tabel 4.13 Jarak Antar Data Iterasi-2.....	91
Tabel 4.14 Jarak Antar Data Iterasi-3.....	91
Tabel 4.15 Jarak Antar Data Iterasi-4.....	92

Tabel 4.16 Jarak Antar Data Iterasi-5.....	93
Tabel 4.17 Jarak Antar Data Iterasi-6.....	93
Tabel 4.18 Hasil Klaster Untuk 2, 3, 4, dan 5 Klaster.....	94
Tabel 4.19 Hasil Klasterisasi Nilai Statistik Uji Untuk Data Set 1 dengan Range Data (1-50) Sebanyak 50 Data Menjadi 2 Klaster	96
Tabel 4.20 Titik Pusat.....	96
Tabel 4.21 Ukuran Kemiripan.....	98
Tabel 4.22 Nilai Index Davies-Bouldin.....	98
Tabel 4.23 Jumlah Klaster Optimal.....	99
Tabel 4.24 Klaster Optimal yang Terbentuk	100
Tabel 5.1 Jumlah Klaster Optimal Beserta Nilai Index Davies-Bouldin	107

DAFTAR SINGKATAN DAN LAMBANG

Simbol	Nama	Pemakaian Pertama Kali pada Halaman
Σ	Sigma.....	12
$f(x)$	Fungsi Kepadatan Peluang.....	15
$F(x)$	Fungsi Distribusi Kumulatif.....	16
β	Beta.....	17
α	Alpha.....	17
ϕ	Phi Variant.....	18
Φ	Phi.....	18
$\Gamma(\alpha)$	Fungsi Gamma.....	19
$L(\theta)$	Fungsi <i>likelihood</i>	27
$l(\theta)$	Fungsi <i>log-likelihood</i>	28
λ	Lambda.....	28

DAFTAR LAMPIRAN

Lampiran 1. Syntax	111
Lampiran 2. Data Simulasi.....	112
Lampiran 3. Tabel Kolmogorov-Smirnov.....	113
Lampiran 4. Tabel Anderson-Darling.....	114