

**ANALISIS METODE K-MEANS DAN FUZZY
C-MEANS CLUSTERING DALAM
MENGELOMPOKKAN KESEJAHTERAAN
MASYARAKAT DITINJAU DARI BIDANG
PENDIDIKAN DI INDONESIA**

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Matematika**







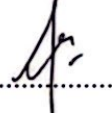

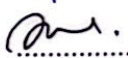

**DINDA TRIANA
1305619026**

**PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA
2023**

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

ANALISIS METODE *K-MEANS* DAN *FUZZY C-MEANS* *CLUSTERING* DALAM MENGELOMPOKKAN KESEJAHTERAAN MASYARAKAT DITINJAU DARI BIDANG PENDIDIKAN DI INDONESIA

Nama : Dinda Triana
No. Registrasi : 1305619026

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N, M.Si.</u> NIP. 196405111989032001		28-08-2023
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		28-08-2023
Ketua	: <u>Drs. Sudarwanto, M.Si., DEA.</u> NIP. 197210262001122001		22-08-2023
Sekretaris	: <u>Dr. Eti Dwi Wiraningsih, S.Pd., M.Si.</u> NIP. 198102032006042001		21-08-2023
Penguji	: <u>Vera Maya Santi, M.Si.</u> NIP. 197905312005012006		23-08-2023
Pembimbing I	: <u>Dr. Yudi Mahatma, M.Si.</u> NIP. 197610202008121001		23-08-2023
Pembimbing II	: <u>Ibnu Hadi, M.Si.</u> NIP. 198107182008011017		23-08-2023

Dinyatakan lulus ujian skripsi tanggal: 16 Agustus 2023

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul " **Analisis Metode K-Means dan Fuzzy C-Means Clustering dalam Mengelompokkan Kesejahteraan Masyarakat Ditinjau dari Bidang Pendidikan di Indonesia**" yang disusun sebagai syarat untuk memperoleh gelar Sarjana Matematika dari Program Studi Matematika Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari Dosen Pembimbing.

Sumber Informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, Agustus 2023

Yang membuat pernyataan



Dinda Triana



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Dinda Triana
NIM : 1305619026
Fakultas/Prodi : FMIPA / Matematika
Alamat email : dinda.irkham19@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Analisis Metode K-Means dan Fuzzy C-Means Clustering dalam
Mengelompokkan Kesejahteraan Masyarakat ditinjau dari Bidang
Pendidikan di Indonesia

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta , 29 Agustus 2023

Penulis

(Dinda Triana)
nama dan tanda tangan

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala rahmat, kebesarannya, pengetahuan, dan kemampuan yang telah diberikan, sehingga penulis mampu menyelesaikan skripsi yang berjudul "Analisis Metode K-Means dan Fuzzy C-Means Clustering dalam Mengelompokkan Kesejahteraan Masyarakat Ditinjau dari Bidang Pendidikan di Indonesia" yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Jurusan Matematika di Universitas Negeri Jakarta.

Skripsi ini berhasil diselesaikan tidak terlepas dari adanya bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih terutama kepada:

1. Bapak Dr. Yudi Mahatma, M.Si selaku Dosen Pembimbing I dan Dosen Pembimbing Akademik dan Bapak Ibnu Hadi, M.Si selaku Dosen Pembimbing II, yang telah meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta arahan sehingga skripsi ini dapat menjadi lebih baik dan terarah.
2. Ibu Lukita Ambarwati, S.Pd., M.Si. selaku Ketua Prodi Matematika FMIPA UNJ yang telah banyak membantu penulis.
3. Seluruh Bapak/Ibu dosen atas pengajarannya yang telah diberikan, serta karyawan/karyawati FMIPA UNJ yang telah memberikan informasi yang penulis butuhkan dalam menyelesaikan skripsi.
4. Papa dan Mama yang selalu mendukung, memberi motivasi, dan setia membantu penulis dengan penuh cinta dan kasih sayang yang tulus.
5. Kak Dede, Kak Dandi, Dozan, dan Dicky yang senantiasa kebersamai penulis dalam menyelesaikan skripsi ini dan terus memberi semangat, mendoakan penulis, serta selalu menghibur ketika penulis mengalami kesulitan.
6. Sahabat karib, Nayla, Dini, Tyna, Yasmin, dan Amel serta teman-teman seperjuangan Matematika 2019 penulis yang selalu ada dikala senang maupun

susah, berproses bersama dan memberikan motivasi agar segera menyelesaikan skripsi ini, juga sebagai teman sepemikiran yang kerap memberikan perhatian kepada penulis.

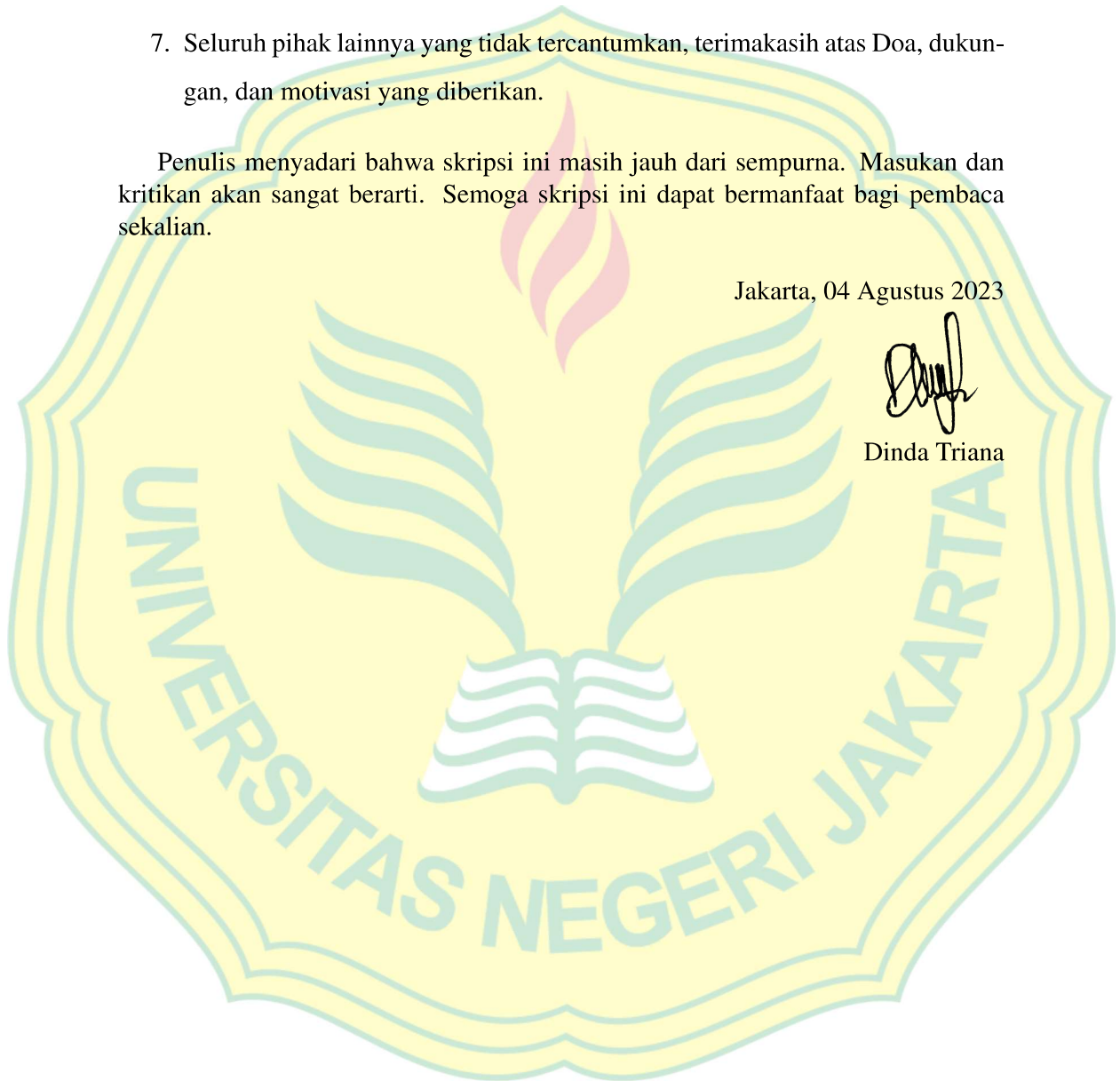
7. Seluruh pihak lainnya yang tidak tercantumkan, terimakasih atas Doa, dukungan, dan motivasi yang diberikan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Masukan dan kritikan akan sangat berarti. Semoga skripsi ini dapat bermanfaat bagi pembaca sekalian.

Jakarta, 04 Agustus 2023



Dinda Triana



ABSTRAK

DINDA TRIANA. Analisis Metode K-Means dan Fuzzy C-Means Clustering dalam Mengelompokkan Kesejahteraan Masyarakat Ditinjau dari Bidang Pendidikan di Indonesia. Skripsi, Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta, Agustus 2023.

Clustering merupakan teknik untuk mengelompokkan data berdasarkan perhitungan jarak. Objek yang memiliki tingkat kehomogenan besar akan berada dalam *cluster* yang sama. Dalam *clustering* terdapat banyak metode diantaranya adalah *K-Means* dan *Fuzzy C-Means* dimana kedua metode tersebut merupakan *clustering* berbasis partisi namun terdapat perbedaan diantara keduanya yaitu, *K-Means* merupakan *Hard Clustering* sedangkan *Fuzzy C-Means* merupakan *Soft Clustering*. Penelitian ini bertujuan untuk menganalisis metode *K-Means* dan *Fuzzy C-Means* dalam mengelompokkan kesejahteraan masyarakat ditinjau dari bidang pendidikan pada 34 provinsi di Indonesia. Kedua Metode dianalisis dengan jumlah *cluster* $K = 2, 3, \dots, 10$ berbantuan *software R Studio*. Pada metode *K-Means* digunakan metode *Elbow* untuk memperoleh *cluster* optimal. Hasil metode *Elbow* menunjukkan bahwa jumlah *cluster* optimal adalah 5 *cluster*. Sedangkan pada metode *Fuzzy C-Means* untuk memperoleh jumlah *cluster* optimal dilihat dari nilai *Dunn Index*. Nilai *Dunn Index* terbesar ada pada 2 *cluster*. Hasil penelitian menunjukkan bahwa perolehan nilai *Davies Bouldin Index* pada metode *K-Means* adalah 1,2675 sedangkan pada metode *Fuzzy C-Means* adalah 2,1359. Pengelompokan kesejahteraan masyarakat ditinjau dari bidang pendidikan pada 34 provinsi di Indonesia lebih baik menggunakan metode *K-Means*.

Kata kunci : *K-Means, Fuzzy C-Means, Davies Bouldin Index, Metode Elbow, Dunn Index.*

ABSTRACT

DINDA TRIANA. Analysis of K-Means and Fuzzy C-Means Methods for Clustering Social Welfare by Education in Indonesia. Undergraduate Thesis, Mathematics Study Program , Faculty of Mathematics and Natural Sciences, State University of Jakarta, August 2023.

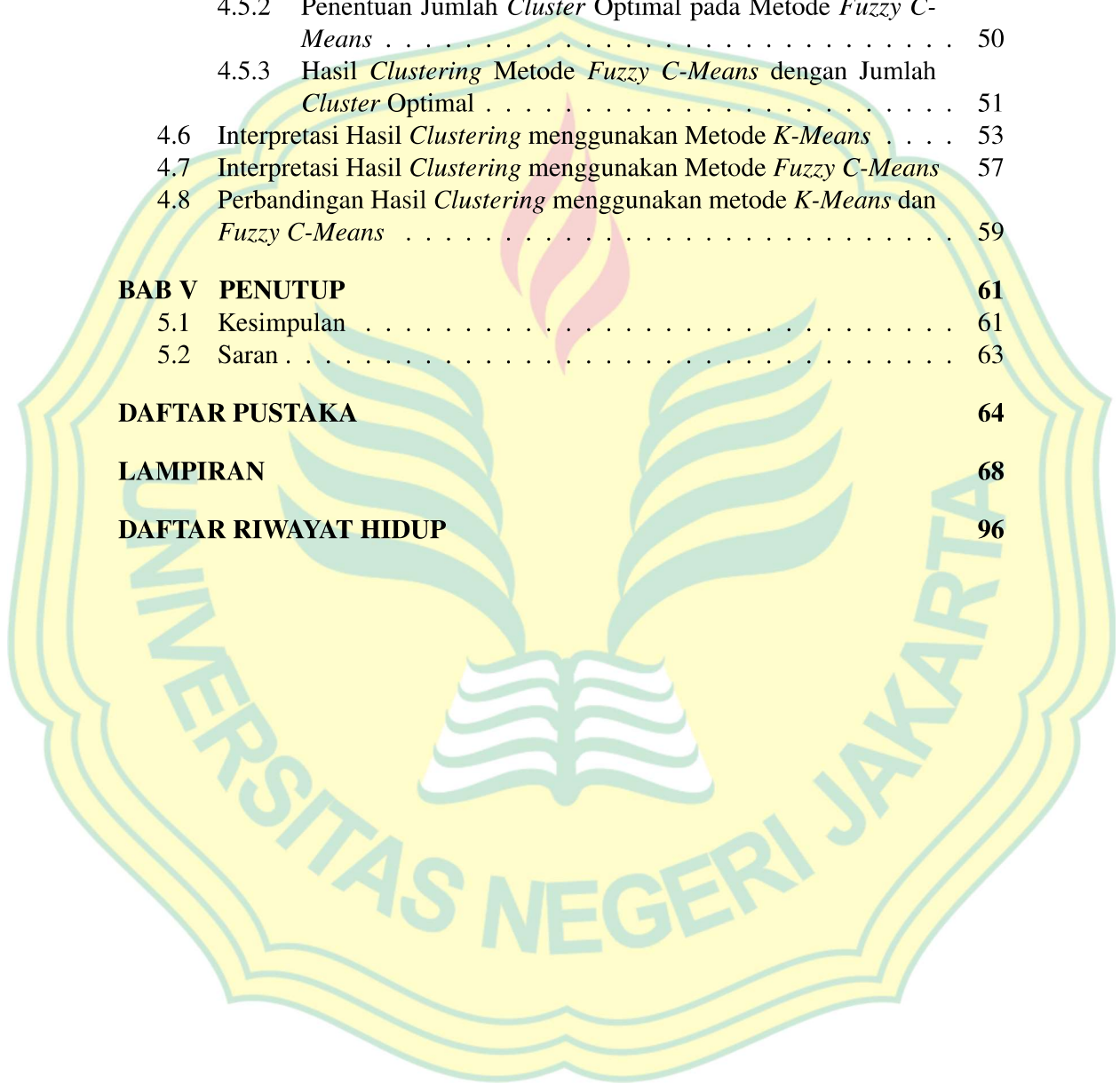
Clustering is a technique for grouping data based on distance. Objects that have a high homogeneous level will be within the same cluster. In clustering, there are many methods including K-Means and Fuzzy C-Means where both methods are partitiofn based clustering but there are differences between which K-Means is Hard Clustering while Fuzzy C-Means is Soft Clustering. This study aims to analyze the K-Means and Fuzzy C-Means Methods for grouping social welfare by education of 34 provinces in Indonesia. Both methods were analyzed with the number of cluster $K = 2, 3, \dots, 10$ using software R Studio. To obtain the optimal cluster at the K-Means method, the Elbow method is used. The results of the Elbow method is that the optimal cluster number is 5 clusters. While in the Fuzzy C-Means method to obtain the optimal number of clusters, the number of Dunn Index is used. The greatest number of Dunn Index is at the number of 2 clusters. The results of the study showed that the acquisition of the value of Davies Bouldin Index in the K-Means method is 1.2675 while the method Fuzzy C-Means is 2.1359. The clustering of social welfare by education in 34 provinces in Indonesia is better to use K-Means Method.

Keywords : *K-Means, Fuzzy C-Means, Davies Bouldin Index, Elbow Method, Dunn Index.*

DAFTAR ISI

DAFTAR ISI	vi
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	6
BAB II LANDASAN TEORI	8
2.1 <i>Data Mining</i>	8
2.2 <i>Clustering</i>	10
2.3 <i>Z-Score Standardization</i>	13
2.4 Ruang Vektor Euclidean	14
2.5 <i>Jarak Square Euclidean</i>	15
2.6 <i>K-Means Clustering</i>	16
2.7 <i>Fuzzy C-Means Clustering</i>	17
2.8 Metode <i>Elbow</i>	19
2.9 Uji Validitas <i>Clustering</i>	22
2.10 Kesejahteraan Masyarakat	24
2.11 Indikator Pendidikan	25
BAB III METODOLOGI PENELITIAN	28
3.1 Data dan Sumber Data	28
3.2 Variabel Penelitian	28
3.3 Langkah Penelitian	30
3.4 Diagram Alir	31
BAB IV PEMBAHASAN	32
4.1 <i>Preprocessing Data</i>	32
4.2 Analisis Statistik Deskriptif	36
4.3 Standarisasi Data	37
4.4 <i>Clustering</i> dengan Metode <i>K-Means</i>	39
4.4.1 Perhitungan Matriks Jarak <i>Euclidean</i>	39
4.4.2 <i>K-Means Clustering</i>	40
4.4.3 Penentuan Jumlah <i>Cluster</i> Optimal pada Metode <i>K-Means</i>	42

4.4.4	Hasil <i>Clustering</i> Metode <i>K-Means</i> dengan Jumlah <i>Cluster</i> Optimal	45
4.5	<i>Clustering</i> dengan Metode <i>Fuzzy C-Means</i>	48
4.5.1	<i>Fuzzy C-Means Clustering</i>	48
4.5.2	Penentuan Jumlah <i>Cluster</i> Optimal pada Metode <i>Fuzzy C-Means</i>	50
4.5.3	Hasil <i>Clustering</i> Metode <i>Fuzzy C-Means</i> dengan Jumlah <i>Cluster</i> Optimal	51
4.6	Interpretasi Hasil <i>Clustering</i> menggunakan Metode <i>K-Means</i>	53
4.7	Interpretasi Hasil <i>Clustering</i> menggunakan Metode <i>Fuzzy C-Means</i>	57
4.8	Perbandingan Hasil <i>Clustering</i> menggunakan metode <i>K-Means</i> dan <i>Fuzzy C-Means</i>	59
BAB V	PENUTUP	61
5.1	Kesimpulan	61
5.2	Saran	63
	DAFTAR PUSTAKA	64
	LAMPIRAN	68
	DAFTAR RIWAYAT HIDUP	96



DAFTAR TABEL

Tabel 2.1	Karakteristik Umum Keempat Metode <i>Clustering</i>	12
Tabel 2.2	Contoh Nilai <i>SSE</i> Metode <i>Elbow</i>	21
Tabel 2.3	Contoh Hasil Perhitungan Sudut Metode <i>Elbow</i>	21
Tabel 3.1	Indikator Pendidikan yang menjadi Variabel Penelitian	28
Tabel 4.1	Variabel Penelitian	32
Tabel 4.2	Provinsi yang Menjadi Objek Penelitian	33
Tabel 4.3	Jumlah Data yang Hilang pada Tiap Variabel	34
Tabel 4.4	Data Penelitian Terproses	35
Tabel 4.5	Statistik Deskriptif	37
Tabel 4.6	Data Terstandarisasi	38
Tabel 4.7	Hasil <i>K-Means Clustering</i> saat $K = 2, 3, \dots, 10$	40
Tabel 4.8	Hasil Akurasi <i>K-Means Clustering</i> pada $K = 2, 3, \dots, 10$	41
Tabel 4.9	Pusat <i>Cluster</i> pada Saat $K = 5$	42
Tabel 4.10	Nilai <i>SSE</i> Metode <i>Elbow</i>	43
Tabel 4.11	Hasil Perhitungan Sudut Metode <i>Elbow</i>	44
Tabel 4.12	<i>Clustering K-Means</i>	45
Tabel 4.13	Nilai <i>SSW</i> dan Akurasi <i>Clustering K-Means</i>	46
Tabel 4.14	Hasil <i>Fuzzy C-Means Clustering</i> saat $K = 2, 3, \dots, 10$	48
Tabel 4.15	Nilai <i>Dunn Index Fuzzy C-Means</i>	50
Tabel 4.16	<i>Clustering dengan Fuzzy C-Means</i>	51
Tabel 4.17	Cluster Provinsi dengan metode <i>K-Means</i>	53
Tabel 4.18	Rata-rata Provinsi <i>K-Means</i>	55
Tabel 4.19	Cluster Provinsi dengan metode <i>Fuzzy C-Means</i>	57
Tabel 4.20	Rata-rata Provinsi <i>Fuzzy C-Means</i>	58
Tabel 4.21	Nilai <i>Davies Bouldin Index</i> Masing-Masing Metode	60

DAFTAR GAMBAR

Gambar 2.1	Tahapan dalam <i>Data Mining</i> pada model CRISP-DM	9
Gambar 2.2	Ilustrasi Clustering	13
Gambar 2.3	Grafik Metode Elbow	21
Gambar 3.1	Diagram Alir Penelitian	31
Gambar 4.1	Hasil <i>Cluster</i> Optimal	44
Gambar 4.2	Plot <i>Clustering K-means</i>	47
Gambar 4.3	Nilai bobot, maxit, dan error dalam <i>Clustering Fuzzy C-Means</i>	52
Gambar 4.4	Plot <i>Clustering Fuzzy C-Means</i>	53

