

**PERBANDINGAN ALGORITMA *K-MEANS* DENGAN *K-MEDOIDS* UNTUK
PENGELOMPOKKAN TINGKAT KRIMINALITAS DI INDONESIA
TAHUN 2020**

*COMPARISON OF K-MEANS AND K-MEDOIDS ALGORITHMS FOR
CRIMINALITY GROUPS IN INDONESIA YEAR 2020*



MUHAMMAD ULIL FAHMI

NIM. 19106010045

**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023

SKRIPSI

**PERBANDINGAN ALGORITMA *K-MEANS* DENGAN *K-MEDOIDS* UNTUK
PENGELOMPOKKAN TINGKAT KRIMINALITAS DI INDONESIA
TAHUN 2020**

*COMPARISON OF K-MEANS AND K-MEDOIDS ALGORITHMS FOR
CRIMINALITY GROUPS IN INDONESIA YEAR 2020*

**Diajukan untuk memenuhi salah satu syarat memperoleh derajat
Sarjana Sains Ilmu Matematika**



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MUHAMMAD ULIL FAHMI

NIM. 19106010045

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Muhammad Ulil Fahmi

NIM : 19106010045

Judul Skripsi : Perbandingan Algoritma *K-Means* dengan *K-Medoids* untuk Pengelompokan Tingkat Kriminalitas di Indonesia Tahun 2020

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

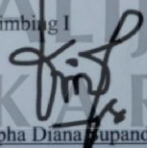
Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 17 Maret 2023

Pembimbing I


Dr. Epha Diana Supandi, S.Si., M.Sc.

NIP. 19750912200801 2 015



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-923/Un.02/DST/PP.00.9/04/2023

Tugas Akhir dengan judul : PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS UNTUK
PENGELOMPOKAN TINGKAT KRIMINALITAS DI INDONESIA TAHUN 2020

yang dipersiapkan dan disusun oleh:

Nama : MUHAMMAD ULIL FAHMI
Nomor Induk Mahasiswa : 19106010045
Telah diujikan pada : Selasa, 28 Maret 2023
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 64270d48350b7



Penguji I
Dr. Muhammad Wakhid Musthofa, S.Si.,
M.Si.
SIGNED

Valid ID: 64280bb3eaf4



Penguji II
Muhamad Zaki Riyanto, S.Si., M.Sc.
SIGNED

Valid ID: 64252d6baec58



Yogyakarta, 28 Maret 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Dr. Dra. Hj. Kharul Wardati, M.Si.
SIGNED

Valid ID: 642840e16e849

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Muhammad Ulil Fahmi
NIM : 19106010045
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 17 Maret 2023



Muhammad Ulil Fahmi

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

**“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.
Dan sesungguhnya Allah tidak akan mengubah keadaan suatu kaum, sebelum
mereka mengubah keadaan diri mereka sendiri”**



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Tugas Akhir ini penulis persembahkan kepada:

Almamater UIN Sunan Kalijaga Yogyakarta tercinta.

Bapak (Alm) dan Ibu yang menjadi inspirasi terbesar penulis.

Terimakasih atas do'a, kasih sayang dan cinta yang selalu mengalir setiap harinya, motivasi, dan nasehat bagi penulis serta selalu memberikan lebih dari apa yang dibutuhkan.

Kedua kakak yang selalu memotivasiku sehingga menjadi semangat bagi penulis.

Terimakasih kepada guruku yang telah memberikan ilmu dan waktu.



**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA**

PRAKATA

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulisan skripsi yang berjudul “Perbandingan Algoritma *K-Means* dengan *K-Medoids* untuk Pengelompokan Tingkat Kriminalitas di Indonesia Tahun 2020” dapat terselesaikan guna memenuhi syarat memperoleh derajat kesarjanaan di Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

Shalawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW. Penulis menyadari penulisan skripsi ini dapat diselesaikan berkat bimbingan, motivasi, dan bantuan semua pihak. Oleh karena itu dengan rendah hati disampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, antara lain:

1. Prof. Dr. Phil Al Makin, M.A., selaku rektor UIN Sunan Kalijaga Yogyakarta.
2. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Muchammad Abrori, S.Si., M.Kom., selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Dr. Sugiyanto, S.Si., ST., M.Si dan Malahayati, S.Si., M.Sc., selaku Dosen Penasehat Akademik mahasiswa matematika Angkatan 2019 UIN Sunan Kalijaga Yogyakarta.
5. Dr. Epha Diana Supandi, S.Si., M.Sc., selaku dosen pembimbing skripsi yang telah meluangkan waktu, tenaga, dan pikiran serta selalu mengarahkan penulis dengan sabar dalam menyelesaikan tugas akhir ini.
6. Bapak/Ibu Dosen serta Staf Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
7. Keluarga besar penulis yang selalu memberikan motivasi dan do'a terbaik untuk penulis.

8. Teman-teman Program Studi Matematika yang telah memberikan motivasi, saran, dan masukan kepada penulis dalam penyelesaian skripsi ini.

Penulis menyadari dalam skripsi ini masih banyak kekurangan, untuk itu penulis mengharap segala kritik dan saran yang membangun dalam perbaikan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi pembaca pada umumnya dan dunia pendidikan pada khususnya.

Yogyakarta, 17 Maret 2023



Muhammad Ulil Fahmi



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN TUGAS AKHIR	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN	v
MOTTO.....	vi
HALAMAN PERSEMBAHAN	vii
PRAKATA	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvii
DAFTAR SIMBOL	xviii
INTISARI.....	xxi
ABSTRACT	xxii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan Masalah.....	4
1.3 Rumusan Masalah	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Keaslian Penelitian	5
1.7 Tinjauan Pustaka	5
1.8 Sistematika Penulisan	9
BAB II LANDASAN TEORI	10
2.1 Matriks	10

2.1.1	Jenis Matriks	11
2.1.2	Penjumlahan Matriks	11
2.1.3	Perkalian Matriks dengan Skalar	12
2.1.4	Perkalian Dua Matriks	12
2.1.5	Transpose Matriks	13
2.1.6	Determinan Matriks	14
2.1.7	Invers Matriks	14
2.2	Statistika	15
2.2.1	Rata-rata	15
2.2.2	Variansi	15
2.2.3	Variansi-Kovariansi	16
2.3	Analisis Multivariat	17
2.3.1	Matriks Data Multivariat	19
2.3.2	Matriks Jarak	20
2.3.3	Vektor Mean	21
2.3.4	Matriks Kovariansi	21
2.3.5	Matriks Korelasi	22
2.4	Analisis Korelasi	23
2.4.1	Korelasi <i>Product Moment Pearson</i>	24
2.4.2	Korelasi <i>Rank Spearman</i>	25
2.4.3	Korelasi <i>Kendall Tau</i>	26
2.5	Distribusi	26
2.5.1	Distribusi Normal	26
2.5.2	Distribusi Normal Multivariat	26
2.6	Data Mining	27
2.6.1	Tahapan Data Mining	30
2.7	<i>Clustering</i>	32

2.7.1	Konsep Dasar <i>Clustering</i>	32
2.7.2	Metode <i>Clustering</i>	34
2.7.3	Uji Asumsi	36
2.7.4	Standarisasi Data	37
2.7.5	<i>Outlier</i>	38
2.7.6	<i>Mahalanobis Distance</i>	38
2.7.7	Evaluasi <i>Cluster</i> dengan DBI	39
2.8	Analisis Varians (ANOVA)	39
2.8.1	Model ANOVA	39
2.9	Kriminalitas	42
BAB III METODOLOGI PENELITIAN		44
3.1	Studi Pustaka	44
3.2	Jenis Penelitian	44
3.3	Objek, Variabel, Jenis, dan Sumber Data Penelitian	44
3.4	Metodologi Penelitian	45
3.5	Alat Pengolah Data	45
3.6	Skema Penelitian	46
3.7	Pengumpulan Data	47
3.8	Prapemrosesan Data	47
3.8.1	<i>Data Reduction</i>	47
3.8.2	<i>Data Integration</i>	47
3.8.3	<i>Data Cleaning</i>	47
3.8.4	<i>Data Transformation</i>	48
3.9	Timeline Skripsi	49
BAB IV PEMBAHASAN		50
4.1	<i>K-Means</i>	50
4.1.1	Contoh Perhitungan <i>K-Means</i>	52

4.2	<i>K-Medoids</i>	55
4.2.1	Contoh Perhitungan <i>K-Medoids</i>	58
4.3	<i>Davies Bouldin Index</i>	60
4.3.1	Contoh Perhitungan DBI pada <i>K-Means</i>	62
4.3.2	Contoh Perhitungan DBI pada <i>K-Medoids</i>	63
BAB V STUDI KASUS.....		66
5.1	Data Kriminalitas.....	66
5.2	Statistik Deskriptif.....	67
5.3	Uji Asumsi Analisis <i>Cluster</i>	73
5.3.1	Uji Kecukupan Sampel (KMO).....	73
5.3.2	Uji Multikolinearitas.....	73
5.4	Standarisasi Data.....	74
5.5	Prapemrosesan Data.....	74
5.5.1	<i>Data Reduction</i>	74
5.5.2	<i>Data Integration</i>	74
5.5.3	<i>Data Cleaning</i>	74
5.5.4	<i>Data Transformation</i>	76
5.6	<i>K-Means</i> dan <i>K-Medoids</i>	76
5.6.1	<i>K-Means</i>	76
5.6.2	<i>K-Medoids</i>	80
5.7	<i>Davies Bouldin Index</i>	85
BAB VI PENUTUP.....		86
6.1	Kesimpulan.....	86
6.2	Saran.....	88
DAFTAR PUSTAKA.....		89
LAMPIRAN.....		93
DAFTAR RIWAYAT HIDUP.....		112

DAFTAR TABEL

Tabel 1.1.	Tinjauan Pustaka	7
Tabel 2.1.	Matriks Data Multivariat	20
Tabel 2.2.	Contoh Data	33
Tabel 2.3.	Tabel Data	40
Tabel 2.4.	Tabel ANOVA	41
Tabel 3.1.	Timeline Skripsi	49
Tabel 4.1.	Contoh Data <i>K-Means</i>	52
Tabel 4.2.	Iterasi Pertama <i>K-Means</i>	52
Tabel 4.3.	Jarak Data ke <i>Centroid</i> Pertama dan Hasil <i>Cluster</i>	53
Tabel 4.4.	Iterasi Kedua <i>K-Means</i>	53
Tabel 4.5.	Jarak Data ke <i>Centroid</i> Kedua dan Hasil <i>Cluster</i>	54
Tabel 4.6.	Iterasi Ketiga <i>K-Means</i>	54
Tabel 4.7.	Jarak Data ke <i>Centroid</i> Ketiga dan Hasil <i>Cluster</i>	55
Tabel 4.8.	Contoh Data <i>K-Medoids</i>	58
Tabel 4.9.	Iterasi Pertama <i>K-Medoids</i>	58
Tabel 4.10.	Jarak Data ke <i>Medoid</i> Pertama dan Hasil <i>Cluster</i>	59
Tabel 4.11.	Iterasi Kedua <i>K-Medoids</i>	59
Tabel 4.12.	Jarak Data ke <i>Medoid</i> Kedua dan Hasil <i>Cluster</i>	60
Tabel 4.13.	Contoh Data dan Iterasi Ketiga <i>K-Means</i>	62
Tabel 4.14.	Hasil <i>SSW K-Means</i>	62
Tabel 4.15.	Hasil <i>SSB K-Means</i>	63
Tabel 4.16.	Hasil <i>Rmax K-Means</i>	63
Tabel 4.17.	Contoh Data dan Iterasi Pertama <i>K-Medoids</i>	63
Tabel 4.18.	Hasil <i>SSW K-Medoids</i>	64
Tabel 4.19.	Hasil <i>SSB K-Medoids</i>	64

Tabel 4.20.	Hasil R_{max} <i>K-Medoids</i>	64
Tabel 5.1.	Statistik Deskriptif	72
Tabel 5.2.	Uji Kecukupan Sampel.....	73
Tabel 5.3.	Uji Multikolinearitas Pertama.....	73
Tabel 5.4.	Uji Multikolinearitas Kedua	73
Tabel 5.5.	Deteksi <i>Outlier</i>	75
Tabel 5.6.	<i>Cluster</i> Pertama <i>K-Means</i> Tidak Ada <i>Outlier</i>	76
Tabel 5.7.	<i>Cluster</i> Pertama <i>K-Means</i> Ada <i>Outlier</i>	77
Tabel 5.8.	<i>Cluster</i> Kedua <i>K-Means</i> Tidak Ada <i>Outlier</i>	77
Tabel 5.9.	<i>Cluster</i> Kedua <i>K-Means</i> Ada <i>Outlier</i>	78
Tabel 5.10.	<i>Cluster</i> Ketiga <i>K-Means</i> Tidak Ada <i>Outlier</i>	78
Tabel 5.11.	<i>Cluster</i> Ketiga <i>K-Means</i> Ada <i>Outlier</i>	79
Tabel 5.12.	<i>Cluster</i> Pertama <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	81
Tabel 5.13.	<i>Cluster</i> Pertama <i>K-Medoids</i> Ada <i>Outlier</i>	81
Tabel 5.14.	<i>Cluster</i> Kedua <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	82
Tabel 5.15.	<i>Cluster</i> Kedua <i>K-Medoids</i> Ada <i>Outlier</i>	82
Tabel 5.16.	<i>Cluster</i> Ketiga <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	82
Tabel 5.17.	<i>Cluster</i> Ketiga <i>K-Medoids</i> Ada <i>Outlier</i>	83
Tabel 5.18.	Nilai DBI Tidak Ada <i>Outlier</i>	85
Tabel 5.19.	Nilai DBI Ada <i>Outlier</i>	85

DAFTAR GAMBAR

Gambar 2.1. Metode Dependensi.....	18
Gambar 2.2. Analisis Interdependensi	19
Gambar 2.3. Tahapan <i>Data Mining</i>	30
Gambar 2.4. <i>Clustering</i>	32
Gambar 2.5. <i>Hierarki Clustering</i>	35
Gambar 2.6. Proses <i>Clustering</i> dengan Metode <i>K-Means</i>	36
Gambar 2.7. Daerah Penolakan H_0	41
Gambar 3.1. Skema Penelitian.....	46
Gambar 4.1. <i>Flowchart K-Means</i>	51
Gambar 4.2. <i>Flowchart K-Medoids</i>	57
Gambar 5.1. Kejahatan terhadap Nyawa.....	67
Gambar 5.2. Kejahatan terhadap Fisik	67
Gambar 5.3. Kejahatan terhadap Kesusilaan.....	68
Gambar 5.4. Kejahatan terhadap Kemerdekaan Orang.....	69
Gambar 5.5. Kejahatan terhadap Hak Milik.....	69
Gambar 5.6. Kejahatan terkait Narkotika.....	70
Gambar 5.7. Kejahatan terkait Penipuan, Penggelapan, dan Korupsi.....	71
Gambar 5.8. Kejahatan terhadap Ketertiban Umum.....	71
Gambar 5.9. Plot <i>K-Means</i> Tidak Ada <i>Outlier</i>	79
Gambar 5.10. Plot <i>K-Means</i> Ada <i>Outlier</i>	80
Gambar 5.11. Plot <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	83
Gambar 5.12. Plot <i>K-Medoids</i> Ada <i>Outlier</i>	84

DAFTAR LAMPIRAN

Lampiran 1. Data Kriminalitas	93
Lampiran 2. Input Tidak Ada <i>Outlier</i>	94
Lampiran 3. Input Ada <i>Outlier</i>	95
Lampiran 4. Statistik Deskriptif	96
Lampiran 5. Uji Kecukupan Sampel.....	96
Lampiran 6. Uji Multikolinearitas	97
Lampiran 7. Standarisasi Data.....	98
Lampiran 8. Deteksi <i>Outlier</i>	99
Lampiran 9. Menghapus <i>Outlier</i>	100
Lampiran 10. Perhitungan <i>K-Means</i> Tidak Ada <i>Outlier</i>	100
Lampiran 11. Nilai DBI <i>K-Means</i> Tidak Ada <i>Outlier</i>	103
Lampiran 12. Perhitungan <i>K-Means</i> Ada <i>Outlier</i>	103
Lampiran 13. Nilai DBI <i>K-Means</i> Ada <i>Outlier</i>	105
Lampiran 14. Perhitungan <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	106
Lampiran 15. Nilai DBI <i>K-Medoids</i> Tidak Ada <i>Outlier</i>	108
Lampiran 16. Perhitungan <i>K-Medoids</i> Ada <i>Outlier</i>	109
Lampiran 17. Nilai DBI <i>K-Medoids</i> Ada <i>Outlier</i>	111

DAFTAR SIMBOL

<u>Simbol</u>	<u>Arti</u>
(A_{ij})	: elemen pada baris ke- i dan kolom ke- j dalam matriks A
A^t	: transpose A
K_{ij}	: kofaktor dari matriks
A^{-1}	: invers matriks dari A
X_{ij}	: data objek ke- i dan kolom ke- j
n	: banyaknya objek
p	: banyaknya variabel
d_{ij}	: jarak antara objek i dan objek j
x_{ik}	: nilai objek i pada variabel ke k
x_{jk}	: nilai objek j pada variabel ke k
\bar{x}_j	: rata-rata sampel untuk variabel-variabel x_j
x_{ij}	: nilai ke- i dari variabel x_j
μ_j	: rata-rata populasi untuk variabel-variabel x_j
N	: populasi
σ_j^2	: variansi untuk variabel-variabel x_j
S_j^2	: variansi sampel
σ_{jk}	: kovariansi antara dua variabel yaitu x_j dan x_k
μ_k	: rata-rata populasi dari variabel x_k
S_{jk}	: kovariansi antara dua variabel yaitu x_j dan x_k
\bar{x}_k	: rata-rata sampel untuk variabel-variabel x_k
π	: mewakili rasio keliling lingkaran dengan diameternya

∞	: mewakili konsep takhingga
Σ	: lambang operator penjumlahan
r_{ij}	: koefisien korelasi sederhana antara peubah i dan j
ρ_{ij}	: koefisien korelasi parsial antara peubah i dan j
R_k^2	: kuadrat dari koefisien korelasi
Z_{ij}	: data hasil standarisasi observasi ke- i variabel ke- j
s_j	: simpangan baku variabel ke- j
$(x_i - \mu)^T$: vektor transpose dengan $i = 1, 2, \dots, p$
Σ^{-1}	: nilai matriks varians kovarians
d_i^2	: jarak mahalanobis dengan $i = 1, 2, \dots, p$
Y_{ij}	: pengamatan ke- i pada perlakuan ke- j
$T_{.j}$: jumlah pengamatan pada perlakuan ke- j
$T_{..}$: total pengamatan
JKT	: jumlah kuadrat total
JKP	: jumlah kuadrat perlakuan
JKG	: jumlah kuadrat galat
RKP	: rata-rata kuadrat perlakuan
RKG	: rata-rata kuadrat galat
SSW	: <i>sum of square within-cluster</i>
$d(x_j, c_j)$: jarak data dengan <i>centroid</i>
SSB	: <i>sum of square between-cluster</i>
c_i	: <i>cluster</i> satu
c_j	: <i>cluster</i> lainnya
$d(c_i, c_j)$: jarak <i>centroid</i> dengan <i>centroid</i> lainnya

- R : *ratio*
- r_s : koefisien korelasi *rank spearman*
- τ : korelasi *kendall tau*
- r_{xy} : korelasi *pearson*
- $\varepsilon_{i,j}$: komponen *error random*



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

INTISARI

PERBANDINGAN ALGORITMA *K-MEANS* DENGAN *K-MEDOIDS* UNTUK PENGELOMPOKKAN TINGKAT KRIMINALITAS DI INDONESIA TAHUN 2020

Oleh
MUHAMMAD ULIL FAHMI
NIM. 19106010045

Analisis *cluster* merupakan suatu teknik untuk meringkas data, yaitu dengan mengelompokkan objek-objek sesuai dengan karakteristik yang dimiliki masing-masing objek. Metode *K-Means* merupakan suatu metode pengklasteran yang berbasis jarak dimana membagi data ke dalam sejumlah *cluster*. Metode *K-Medoids* merupakan suatu metode pengklasteran dimana metode ini untuk mengatasi kelemahan *K-Means*. Kedua metode tersebut digunakan untuk mengelompokkan data kriminalitas di Indonesia tahun 2020. Pada penelitian ini menggunakan *Davies Bouldin Index* untuk menentukan metode terbaik pada data kriminalitas di Indonesia tahun 2020. *Davies Bouldin Index* (DBI) ini bisa diartikan sebagai semakin kecil nilai DBI maka semakin baik *cluster* yang diperoleh. Hasil penelitian ini menunjukkan terbentuk tiga *cluster* dimana pada metode *K-Means* tidak ada *outlier* menghasilkan nilai DBI sebesar 0,9916 lebih kecil dibandingkan dengan nilai DBI untuk *K-Medoids* tidak ada *outlier* sebesar 1,8733. Sehingga pada kasus ini, metode *K-Means* tidak ada *outlier* merupakan metode yang lebih baik dibandingkan metode *K-Medoids* tidak ada *outlier*.

Kata Kunci: analisis *cluster*, *K-Means*, *K-Medoids*, kriminalitas, dan *Davies Bouldin Index*.

ABSTRACT

COMPARISON OF K-MEANS AND K-MEDOIDS ALGORITHMS FOR CRIMINALITY GROUPS IN INDONESIA YEAR 2020

by

MUHAMMAD ULIL FAHMI

NIM. 19106010045

Cluster analysis is a technique for summarizing data, namely by grouping objects according to the characteristics of each object. The K-Means is a distance-based clustering method which divides data into a number of clusters. K-Medoids method is a clustering method where this method overcomes the weaknesses of K-Means. The second method is used to classify crime data in Indonesia in 2020. In this study, the Davies Bouldin Index was used to determine the best method for data on crime in Indonesia in 2020. The Davies Bouldin Index (DBI) can be interpreted as the smaller the DBI value, the better the cluster obtained. The results of this study indicate that three clusters are formed where the K-Means no outlier method produces a DBI value of 0,9916 which is smaller than the DBI value for K-Medoids no outlier of 1,8733. So in this case, the K-Means no outlier method is a better method than the K-Medoids no outlier method.

Keywords: cluster analysis, K-Means, K-Medoids, crime, and Davies Bouldin Index.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB I

PENDAHULUAN

Pada BAB I ini, membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, tinjauan pustaka, dan sistematika penulisan.

1.1 Latar Belakang

Analisis multivariat merupakan salah satu teknik statistika yang digunakan untuk menganalisis dimana variabelnya lebih dari satu yang dilakukan secara bersamaan. Sehingga dengan teknik analisis ini dapat menganalisis beberapa variabel terhadap variabel lainnya dalam waktu bersamaan. Dalam analisis data multivariat terdapat dua klasifikasi, yaitu metode *dependensi* dan metode *interdependensi*. Dalam analisis statistika multivariat juga terdapat beberapa metode, diantaranya analisis komponen utama, analisis faktor, analisis korelasi kanonik, analisis regresi multivariat, analisis varians multivariat, analisis diskriminan, dan analisis *cluster*.

Pada penelitian ini menggunakan salah satu metode analisis statistika multivariat, yaitu dengan analisis *cluster*. Analisis *cluster* merupakan suatu teknik untuk meringkas data, yaitu dengan mengelompokkan objek-objek sesuai dengan karakteristik yang dimiliki masing-masing objek. Di antara banyaknya analisis *cluster*, disini penulis menggunakan *K-Means clustering* dan *K-Medoids clustering*. Alasan menggunakan metode *K-Means* dan *K-Medoids* adalah karena cepat dan akurasi cukup tinggi untuk memproses objek dalam jumlah yang besar.

Indonesia merupakan suatu negara yang memiliki jumlah penduduk sangat besar. Indonesia menduduki negara dengan jumlah penduduk terbesar di Asia Tenggara. Indonesia juga menduduki negara dengan jumlah penduduk terbesar ke empat di dunia. Pada tahun 2020, Indonesia dengan jumlah penduduk 273,5 juta jiwa. Dengan angka jumlah penduduk sebesar itu, maka terjadinya kepadatan penduduk. Kepadatan penduduk tersebut secara tidak langsung mempengaruhi tingkat terjadinya kriminalitas pada suatu daerah. Di sisi lain, Indonesia juga kekurangan tempat lapangan

pekerjaan dan masih banyak warga yang kurang mampu. Dengan keadaan tersebut, juga mempengaruhi tingkat terjadinya kriminalitas. Di Indonesia pada tahun 2020 terdapat 203.584 kasus tingkat terjadinya kriminalitas. Data tersebut diambil di Badan Pusat Statistik (BPS).

Adanya jumlah kasus kriminalitas sebesar 203.584 menjadi suatu permasalahan yang dialami oleh masyarakat luas. Tindakan kriminalitas itu terdapat berbagai macam jenis. Terjadinya suatu tindakan kriminalitas berada di berbagai wilayah dan waktu yang berbeda sehingga sulit dalam memetakan provinsi yang memiliki tingkat kerawanan kriminalitas. Karena itu, dibutuhkan informasi mengenai pemetaan wilayah sesuai dengan tingkat terjadinya tindakan kriminalitas. Informasi tersebut sangat dibutuhkan oleh masyarakat dan pihak kepolisian. Dimana informasi itu sangat berguna bagi masyarakat untuk lebih waspada saat berada di wilayah terjadinya tindakan kriminalitas. Di sisi lain, pihak kepolisian juga sangat terbantu sehingga dapat menentukan wilayah mana saja yang perlu pengawasan ekstra atau tidak. Pengolahan data tersebut menggunakan teknik *data mining*.

Menurut Buulolo (2020) *data mining* merupakan suatu aktivitas yang berhubungan dengan pengumpulan data, pemakaian data historis dalam menemukan pengetahuan, informasi, keteraturan, pola atau hubungan yang memiliki ukuran data besar. Dapat disimpulkan bahwa *data mining* merupakan proses dimana untuk mendapatkan suatu informasi baru dari data yang berukuran besar sehingga dapat digunakan dalam membantu pengambilan keputusan.

Menurut Grossman dan Frieder (2004) Algoritma *K-Means* merupakan suatu metode *clustering* yang paling sederhana. *K-Means* merupakan metode analisis *cluster non-hierarki* yang mencoba untuk mempartisi data yang ada objek menjadi satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya. Untuk objek yang karakteristik sama dikelompokkan ke dalam *cluster* yang sama sedangkan objek berkarakteristik berbeda dikelompokkan ke dalam *cluster* yang lain. Adapun tujuan pengklasteran data ini adalah untuk meminimalkan fungsi objektif. Sedangkan menurut Vercellis (2009) Algoritma PAM (*Partitioning Around Medoids*) atau biasa juga disebut

dengan algoritma *K-Medoids* merupakan algoritma yang diwakili oleh *cluster* yaitu *medoid*. Perbedaan antara algoritma *K-Medoids* dengan algoritma *K-Means* yaitu algoritma *K-Medoids* menggunakan objek sebagai perwakilan (*medoid*) pusat *cluster* untuk tiap *cluster*, sementara algoritma *K-Means* membutuhkan nilai rata-rata (*mean*) sebagai pusat *cluster*. Menurut Harahap (2021) *Davies-Bouldin Index* (DBI) adalah suatu ukuran yang digunakan untuk menentukan jumlah *cluster* yang terbaik setelah proses pengklasteran selesai. DBI ini memiliki tujuan untuk memaksimalkan jarak antara *cluster* yang satu dengan *cluster* yang lain. Dan juga mencoba meminimalkan jarak antar objek dalam suatu *cluster*. Semakin kecil nilai DBI maka semakin baik *cluster* yang diperoleh.

Dari penelitian Qomariyah (2021) yang berjudul “*Studi Komparasi Algoritma K-Means Clustering dan K-Medoids Clustering Dalam Pengklasteran Data Mahasiswa (Studi Kasus : Data Mahasiswa Fakultas Sains dan Teknologi UIN Sunan Kalijaga)*”. Hasil dari penelitian ini berupa implementasi *K-Means* dan *K-Medoids* untuk menentukan *cluster* peta persebaran wilayah mahasiswa. Dan berdasarkan uji validasi dengan nilai *Davies Bouldin Index* (DBI) bahwa *K-Means* lebih optimal daripada *K-Medoids*. Pada skripsi tersebut juga menggunakan bahasa pemrograman *Python* dan *RapidMiner*.

Pada penelitian ini, penulis menggunakan data kriminalitas. Data tersebut diambil dari BPS pada tahun 2020. Atribut data yang digunakan adalah kejahatan terhadap nyawa, kejahatan terhadap fisik, kejahatan terhadap kesusilaan, kejahatan terhadap kemerdekaan orang, kejahatan terhadap hak milik, kejahatan terkait narkoba, kejahatan terkait (penipuan, penggelapan, dan korupsi), dan kejahatan terhadap ketertiban umum.

Berdasarkan latar belakang yang telah dijelaskan, maka penulis mengambil judul “*Perbandingan Algoritma K-Means dengan K-Medoids untuk Pengelompokan Tingkat Kriminalitas di Indonesia Tahun 2020*”. Pada kasus ini menggunakan metode *K-Means* dan *K-Medoids* bertujuan untuk membandingkan hasil

antara kedua metode tersebut. Dengan hasil kedua metode tersebut maka diperoleh metode terbaik untuk pengelompokkan tingkat terjadinya kriminalitas di Indonesia.

1.2 Batasan Masalah

Pada penulisan ini, penulis membatasi dengan menggunakan dua metode yaitu *K-Means* dan *K-Medoids*. Sedangkan data yang digunakan merupakan data kriminalitas di Indonesia pada tahun 2020. Datanya mengambil dari BPS (Badan Pusat Statistik).

1.3 Rumusan Masalah

1. Bagaimana hasil pengklasteran pada data kriminalitas di Indonesia tahun 2020?
2. Bagaimana hasil perbandingan pengklasteran data kriminalitas pada tahun 2020 dengan metode *K-Means* dan *K-Medoids*?
3. Bagaimana hasil dalam menentukan metode terbaik pada data kriminalitas di Indonesia tahun 2020 dengan menggunakan *Davies Bouldin Index*?

1.4 Tujuan Penelitian

1. Mengetahui hasil pengklasteran pada data kriminalitas di Indonesia tahun 2020
2. Mengetahui hasil perbandingan pengklasteran data kriminalitas pada tahun 2020 dengan metode *K-Means* dan *K-Medoids*
3. Mengetahui hasil dalam menentukan metode terbaik pada data kriminalitas di Indonesia tahun 2020 dengan menggunakan *Davies Bouldin Index*

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini, sebagai berikut:

1. Bagi Peneliti
 - a. Menambah pengetahuan tentang algoritma *K-Means* dan *K-Medoids*
 - b. Mengatahui bagaimana menyelesaikan suatu data kriminalitas dengan menggunakan algoritma *K-Means* dan *K-Medoids*
 - c. Melatih keuletan, kesabaran dan ketelitian dalam menganalisis suatu data

2. Bagi Pembaca

Sebagai informasi mengenai pemetaan wilayah sesuai tingkat terjadinya kriminalitas sehingga lebih waspada jika sedang berada di wilayah tersebut. Bisa juga dijadikan suatu referensi untuk penelitian selanjutnya. Dan untuk penelitian ini dapat memberikan suatu pemikiran baru untuk pembaca serta bisa teruskan dan juga bisa dikembangkan.

3. Bagi Kepolisian

Pada penelitian ini bisa dimanfaatkan oleh kepolisian. Dimana dengan penelitian ini dapat membantu kepolisian dalam hal pengawasan yang lebih atau tidak pada daerah tingkat terjadinya kriminalitas

1.6 Keaslian Penelitian

Penelitian mengenai perbandingan algoritma *K-Means* dan *K-Medoids* pada data kriminalitas di Indoensia sepengetahuan penulis belum pernah dilakukan sebelumnya. Untuk komparasi algoritma *K-Means* dan *K-Medoids* seperti ini pernah dilakukan sebelumnya tetapi perbedaannya pada studi kasus yang diteliti, bahasa pemrogramannya, proses perhitungan, dan data-data yang lain.

1.7 Tinjauan Pustaka

Pada penelitian ini, dimaksudkan bagi peneliti dalam menggunakan beberapa tinjauan pustaka, yaitu jurnal, buku, skripsi, dan lain-lain yang berkaitan dengan penelitian ini. Tinjauan pustaka sebagai berikut:

1. Skripsi tentang “*Perbandingan Algoritma K-Means, X-Means, dan K-Medoids untuk Klasterisasi Awak Kabin pada PT. Lion Mentari Airlines*” Oleh Ahmad Jurnaidi Wahidin mahasiswa S2 Universitas Budi Luhur. Dalam skripsi tersebut membahas tentang Algoritma *K-Means*, *X-Means*, dan *K-Medoids* dengan membandingkannya pada awak kabin di PT. Lion Mentari Airlines. Selanjutnya, pencarian algoritma terbaik dengan metode DBI. Kemudian algoritma terbaik itu diimplementasikan ke dalam program dan dapat diterima oleh calon user yang

berdasarkan hasil *Software Quality Assurance* dan hasil pengujian akurasi program.

2. Jurnal tentang “*Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas*” Oleh Aceng Supriyadi, Agung Triayudi, dan Ira Diana Sholihati. Dalam jurnal ini membahas tentang perbandingan algoritma *K-Means* dan *K-Medoids* yang dilakukan pada pengelompokan armada kendaraan truk berdasarkan produktivitas. Dimana dilakukan perbandingan menggunakan *Davies Bouldin Index*. Setelah itu, dengan algoritma terbaik diimplementasikan pada pembuatan aplikasi clustering armada kendaraan berbasis web.
3. Jurnal tentang “*Perbandingan Metode Fuzzy C-Means dan K-Means untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang*” Oleh Hana Sugiastu Firdaus, Arief Laila Nugraha, Bandi Sasmito, Moehammad Awaluddin, dan Chairunisa Afnidya Nanda. Dalam jurnal ini membahas tentang memetakan daerah rawan kriminalitas di kota Semarang menggunakan metode *Fuzzy C-Means* dan *K-Means*. Dilakukan perbandingan antara kedua metode tersebut. Dimana metode *Fuzzy C-Means* menggunakan *Partition Coefficient Index* sedangkan metode *K-Means* menggunakan *Silhouette Indeks*.
4. Skripsi tentang “*Studi Komparasi Algoritma K-Means Clustering dan K-Medoids Clustering Dalam Pengklasteran Data Mahasiswa (Studi Kasus : Data Mahasiswa Fakultas Sains dan Teknologi UIN Sunan Kalijaga)*” Oleh Qomariyah mahasiswa S1 UIN Sunan Kalijaga Yogyakarta. Dalam skripsi tersebut membahas tentang algoritma *K-Means* dan *K-Medoids* dimana membandingkannya pada data mahasiswa fakultas sains dan teknologi. Berdasarkan uji validasi dengan nilai *Davies Bouldin Index* (DBI) bahwa *K-Means* lebih optimal daripada *K-Medoids*.
5. Jurnal tentang “*Pengklasteran Kabupaten/Kota di Jawa Tengah berdasarkan Tenaga Kesehatan dengan Menggunakan Metode Ward dan K-Means*” Oleh Sri Puji Lestari, Epha Diana Supandi, dan Pipit Pratiwi Rahayu. Dalam jurnal ini membahas tentang pengklasteran kabupaten/kota di Jawa Tengah berdasarkan

tenaga kesehatan. Dimana membandingkan metode *Ward* dan *K-Means*. Berdasarkan dengan nilai rasio simpangan baku menunjukkan bahwa metode *K-Means* lebih baik daripada metode *Ward*.

Tabel 1.1. Tinjauan Pustaka

No	Peneliti (Tahun)	Objek Peneliti	Metode	Hasil
1.	Ahmad Jurnaidi Wahidin (2018)	awak kabin di PT. Lion Mentari Airlines	<i>K-Means</i> , <i>X-Means</i> , dan <i>K-Medoids</i>	Algoritma <i>K-Means</i> menjadi algoritma terbaik untuk klusterisasi awak kabin di PT. Lion Mentari Airlines daripada <i>X-Means</i> dan <i>K-Medoids</i> .
2.	Aceng Supriyadi, Agung Triayudi, dan Ira Diana Sholihati (2021)	armada kendaraan truk berdasarkan produktivitas	<i>K-Means</i> dan <i>K-Medoids</i>	Algoritma <i>K-Means</i> lebih optimal daripada <i>K-Medoids</i> .

No	Peneliti (Tahun)	Objek Peneliti	Metode	Hasil
3	Hana Sugiastu Firdaus, Arief Laila Nugraha, Bandi Sasmito, Moehammad Awaluddin, dan Chairunisa Afnidya Nanda (2021)	Tindak kriminalitas di Kota Semarang	<i>Fuzzy C-Means</i> dan <i>K-Means</i>	Berdasarkan dengan uji verifikasi maka <i>Fuzzy C-Means</i> memberikan hasil yang lebih baik daripada <i>K-Means</i> .
4	Qomariyah (2021)	Mahasiswa fakultas sains dan teknologi UIN Sunan Kalijaga	<i>K-Means</i> dan <i>K-Medoids</i>	Berdasarkan uji validitas dengan DBI maka didapatkan <i>K-Means</i> lebih optimal daripada <i>K-Medoids</i> .
5	Sri Puji Lestari, Epha Diana Supandi, dan Pipit Pratiwi Rahayu (2018)	Tenaga kesehatan	<i>Ward</i> dan <i>K-Means</i>	Berdasarkan dengan nilai rasio simpangan baku menunjukkan bahwa metode <i>K-Means</i> lebih baik daripada metode <i>Ward</i> .

1.8 Sistematika Penulisan

Yang dimaksud dalam sistematika penulisan ini supaya pembaca lebih mudah dalam memahami tulisan penelitian. Sistematika penulisan dalam penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada BAB I ini, membahas tentang latar belakang, rumusan masalah , batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, tinjauan pustaka, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada BAB II ini, membahas tentang teori-teori pendukung penelitian, meliputi matriks, statistika, analisis multivariat, analisis korelasi, distribusi, *data mining*, *clustering*, dan analisis varians (ANOVA).

BAB III : METODOLOGI PENELITIAN

Pada BAB III ini, membahas tentang langkah-langkah penulis dalam mencapai tujuan penelitian dan mendapatkan kesimpulan akhir dari penelitian.

BAB IV : PEMBAHASAN

Pada BAB IV ini, membahas tentang metode yang dipakai untuk penelitian.

BAB V : STUDI KASUS

Pada BAB V ini, membahas tentang studi kasus yang diambil oleh peneliti, yaitu data kriminalitas. Dan menjelaskan hasil penelitiannya.

BAB VI : PENUTUP

Pada BAB VI ini, membahas tentang kesimpulan yang didapat dari hasil penelitian yang dilakukan oleh peneliti. Pada bab ini juga terdapat saran untuk penelitian selanjutnya.

BAB VI

PENUTUP

Pada BAB VI ini, membahas tentang kesimpulan yang didapat dari hasil penelitian yang dilakukan oleh peneliti. Pada bab ini juga terdapat saran untuk penelitian selanjutnya.

6.1 Kesimpulan

Berdasarkan hasil dan pembahasan maka dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan analisis *cluster* pada data kriminalitas di Indonesia tahun 2020 sebagai berikut:
 - a. Melakukan pendeskripsian data pada semua variabel
 - b. Melakukan uji kecukupan sampel menggunakan uji KMO. Apabila diperoleh masing-masing variabel lebih dari 0.5 dinyatakan bahwa sampel yang diteliti dapat mewakili populasi yang ada atau bisa dikatakan sampel layak digunakan.
 - c. Melakukan uji multikolinearitas dengan kriteria VIF, apabila VIF kurang dari 10 maka tidak terjadi multikolinearitas sehingga variabel dapat digunakan.
 - d. Melakukan standarisasi data, dimana diperlukan untuk rentang nilai pengukuran tidak jauh berbeda dan perhitungan jarak menjadi efektif.
 - e. Melakukan prapemrosesan data:
 - 1) *Data reduction*
Dimana suatu proses dilakukan untuk menghilangkan data atau mengurangi data.

2) *Data integration*

Dimana suatu proses penggabungan dari berbagai database ke dalam satu database yang baru.

3) *Data cleaning*

Dimana suatu proses untuk menghilangkan noise, data yang tidak konsisten, mengisi data *missing value*. Seperti halnya kosongnya data, data duplikasi, data kurang huruf, salah ketik, dan lain-lain. Data yang tidak relevan juga lebih baik untuk dibuang.

4) *Data transformation*

Dimana suatu proses untuk mengubah data ke dalam bentuk format yang sesuai untuk diproses dalam data mining.

- f. Melakukan perhitungan *K-Means* dan *K-Medoids* dimana pada analisis *cluster* ini menggunakan kedua metode tersebut.
 - g. Melakukan perbandingan *K-Means* dan *K-Medoids* dengan menggunakan *Davies Bouldin Index*.
2. Berdasarkan hasil *K-Means* tidak ada *outlier* dan *K-Medoids* tidak ada *outlier* pada data kriminalitas di Indonesia tahun 2020 dapat dijelaskan bahwa kedua metode tersebut sama-sama menghasilkan tiga *cluster*. Dimana pada metode *K-Means* tiga *cluster*, yaitu *cluster* pertama (tinggi), *cluster* kedua (rendah), dan *cluster* ketiga (sedang). Sedangkan pada metode *K-Medoids* tiga *cluster*, yaitu *cluster* pertama (sedang), *cluster* kedua (tinggi), dan *cluster* ketiga (rendah).
 3. Berdasarkan hasil *K-Means* ada *outlier* dan *K-Medoids* ada *outlier* dapat dijelaskan bahwa kedua metode tersebut sama-sama menghasilkan tiga *cluster*. Dimana pada metode *K-Means* tiga *cluster*, yaitu *cluster* pertama (tinggi), *cluster* kedua (rendah), dan *cluster* ketiga (sedang). Sedangkan pada metode *K-Medoids* tiga *cluster*, yaitu *cluster* pertama (rendah), *cluster* kedua (tinggi), dan *cluster* ketiga (sedang).

4. Berdasarkan *Davies Bouldin Index* menunjukkan metode *K-Means* tidak ada *outlier* lebih baik daripada metode *K-Medoids* tidak ada *outlier*. Dikarenakan pada nilai DBI untuk *K-Means* tidak ada *outlier* lebih kecil daripada *K-Medoids* tidak ada *outlier*. Dimana nilai DBI untuk *K-Means* tidak ada *outlier* sebesar 0,9916 dan nilai DBI untuk *K-Medoids* tidak ada *outlier* sebesar 1,8733. Dengan kata lain, metode *K-Means* tidak ada *outlier* lebih baik dalam mengklasterkan data kriminalitas di Indonesia tahun 2020 dibandingkan metode *K-Medoids* tidak ada *outlier*.
5. Berdasarkan *Davies Bouldin Index* menunjukkan metode *K-Medoids* ada *outlier* lebih baik daripada metode *K-Means* ada *outlier*. Dikarenakan pada nilai DBI untuk *K-Medoids* ada *outlier* lebih kecil daripada *K-Means* ada *outlier*. Dimana nilai DBI untuk *K-Medoids* ada *outlier* sebesar 0,9559 dan nilai DBI untuk *K-Means* ada *outlier* sebesar 1,1859. Dengan kata lain, metode *K-Medoids* ada *outlier* lebih baik dalam mengklasterkan data kriminalitas di Indonesia tahun 2020 dibandingkan metode *K-Means* ada *outlier*.

6.2 Saran

Berdasarkan hasil penelitian maka penulis bermaksud memberikan saran yang mudah-mudahan dapat bermanfaat bagi penelitian selanjutnya sebagai berikut:

1. Peneliti selanjutnya diharapkan untuk mengkaji lebih banyak sumber yang terkait dengan analisis *cluster* agar dapat menguasai secara menyeluruh.
2. Peneliti selanjutnya diharapkan menggunakan metode selain *K-Means* dan *K-Medoids*.
3. Peneliti selanjutnya diharapkan menambahkan variabel agar lebih banyak.
4. Peneliti selanjutnya diharapkan menggunakan perbandingan selain *Davies Bouldin Index*.

DAFTAR PUSTAKA

- Agung L. 2021. *Perbandingan Algoritma K-Means dan K-Medoids dalam Penclustering Data Penjualan PT. United Teknologi Integrasi*. Skripsi. Jakarta: Universitas Mercu Buana.
- Ahmad Jurnaidi W. 2018. *Perbandingan Algoritma K-Means, X-Means, dan K-Medoids untuk Klasterisasi Awak Kabin pada PT. Lion Mentari Airlines*. Skripsi. Jakarta: Universitas Budi Luhur.
- Anton, Howard dan Rorres, C. 2004. *Aljabar Linear Elementer Edisi ke-8*. Jakarta: Erlangga.
- Anton, Howard. 2005. *Aljabar Linear Elementer*. Jakarta: Erlangga.
- Butsianto, S. dan Nindi, T.M. 2020. *Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering, Jurnal Nas. Komputasi dan Teknol*, vol. 3, no. 3, hal. 187-201.
- Buulolo, Efori. 2020. *Data Mining untuk Perguruan Tinggi*. Yogyakarta: DeePublish.
- Dewi, Riani Sinta, dkk. 2020. *Klasifikasi Usaha Industri di Kabupaten Luwu Utara Berdasarkan Indikator Industri Kecil dan Industri Menengah Menggunakan Metode Average Linkage Clustering, Journal MSA*, vol. 8, no. 2, hal. 14-19.
- Dewi, Sinta Maulina, dkk. 2019. *Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah, Jurnal SENSASI*. <https://prosiding.seminar-id.com/index.php/sensasi/article/download/376/368>
- Farahdinna, Frenda, dkk. 2019. *Perbandingan Algoritma K-Means dan K-Medoids dalam Klasterisasi Produk Asuransi Perusahaan Nasional, Jurnal Ilmiah Fifo*, vol. XI, no. 2, hal. 208-214.
- Firdaus, Hana Sugiastu, dkk. 2021. *Perbandingan Metode Fuzzy C-Means dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang, Jurnal Geodesi dan Geomatika*, vol. 04, no. 01, hal. 58-64.
- Grossman, D.A. dan Frieder, O. 2004. *Information Retrieval Algorithms and Heuristics Second Edition*. Netherlands: Springer.
- Hair, J.F. et al. 2010. *Multivariate Data Analysis: A Global Perspective. 7th Edition*. New Jersey: Pearson Prentice Hall.
- Han, J., & Kamber, M. 2012. *Data Mining: Concept and Techniques*. Waltham: Morgan Kauffman Publisher.

- Harahap, Fitriana. 2021. *Perbandingan Algoritma K Means dan K Medoids untuk Cluster Siswa Tunagrahita, Terapan Informatika Nusantara*, vol. 2, no. 4, hal. 191-197.
- Irwansyah, Edy dan Muhammad Faisal. 2015. *Advanced Clustering: Teori dan Aplikasi*. Yogyakarta: DeePublish.
- Johnson, R.A. & D.W. Wichern. 2002. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th edition. New Jersey: Prentice Hall.
- Kamila, I., Khairunnisa, U., dan Mustakim. 2019. *Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*.
- Kusrini, E. T. L., & Taufiq, E. 2009. *Algoritma data mining*. Yogyakarta: Andi Offset.
- Lestari, Sri Puji, dkk. 2018. *Pengklasteran Kabupaten/Kota di Jawa Tengah berdasarkan Tenaga Kesehatan dengan Menggunakan Metode Ward dan K-Means, Jurnal Fourier*, vol 7, no. 2, hal. 103-109.
- Lumenta, Christian Y, dkk. 2012. *Analisis Jalur Faktor-Faktor Penyebab Kriminalitas di Kota Manado, Jurnal Ilmiah Sains*, vol. 12, no. 2, hal.77-83.
- Mahalanobis PC. 1936. *On the Generalised Distance in Statistics. Proceedings of the National Institute of Science India*, 2:4955.
- Mustafid. 2003. *Statistika Elementer Metode dan Aplikasi dengan SPSS*. Semarang: Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Diponegoro.
- Neni Nistiyana. 2021. *Penerapan Analisis Faktor dengan Confirmatory Factor Analysis dan Analisis Diskriminan dengan Multiple Discriminant Analysis (Studi Kasus: Faktor-Faktor yang Mempengaruhi Tingkat Kekhawatiran Masyarakat Kabupaten Sleman Yogyakarta terhadap Kondisi Pandemi Covid-19)*. Skripsi. Yogyakarta: UIN Sunan Kalijaga.
- Pramesti, Dyang Falila, dkk. 2017. *Implementasi Metode K-Medoids Clustering untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot), Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, hal.723-732.

- Qomariyah. 2021. *Studi Komparasi Algoritma K-Means Clustering dan K-Medoids Clustering Dalam Pengklasteran Data Mahasiswa (Studi Kasus : Data Mahasiswa Fakultas Sains dan Teknologi UIN Sunan Kalijaga)*. Skripsi. Yogyakarta: UIN Sunan Kalijaga.
- Quadratullah, Muhammad Farhan. 2012. *Analisis Regresi Terapan: Teori, Contoh Kasus, dan Aplikasi dengan SPSS*. Yogyakarta: Andi.
- Ramageri, Bharati M. 2017. *Data Mining Techniques And Applications. Indian Journal of Computer Science and Engineering*, vol. 1, no. 4, hal. 301-305.
- Ramdhani, Fitra, dkk. 2015. *Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster. Jurnal Gaussian*, vol. 4, no. 4, hal. 875-884.
- Rencher, A.C. 2002. *Methods of Multivariate Analysis*. Amerika: John Wiley & Sons, Inc.
- Riski Yuli Artanti. 2018. *Perbandingan Kinerja Metode Average Linkage dan Metode Ward Dalam Menentukan Hasil Analisis Cluster (Studi Kasus: Data Jumlah Kasus Penyakit pada 12 Kecamatan di Kabupaten Kulon Progo Tahun 2015)*. Skripsi. Yogyakarta: UIN Sunan Kalijaga.
- Sarwono, J. 2013. *Statistik Multivariat Aplikasi untuk Riset Skripsi*. Yogyakarta: Penerbit ANDI.
- Setiawan, Sony Sunaryo dan Tiodora Hadumaon Siagian. 2011. *Mengatasi Masalah Multikolinearitas dan Outlier dengan Pendekatan ROBPCA (Studi Kasus Analisis Regresi Angka Kematian Bayi di Jawa Timur, Jurnal Matematika, Saint dan Teknologi*, vol. 12, no. 1, hal. 1-10.
- Simamora, B. 2005. *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka Utama.
- Sugiyono. 2007. *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- Supandi, Epha Diana. 2020. *Statistika dan Terapannya*. Bandung: PT Refika Aditama.
- Supriyadi, Aceng, dkk. 2021. *Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas, Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, vol. 06, no. 02, hal. 229-240.
- Tan, P.N., Steinbach, M., dan Kumar, V., 2006. *Introduction to Data Mining*. Boston: Pearson Education.

- Vandanu Aditya Putra S. 2018. *Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015*. Skripsi. Yogyakarta: Universitas Islam Indonesia.
- Vercillis, Carlo. 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. Milan: WILEY.
- Wustqa, Dhoriva Urwatul, dkk. 2018. *Analisis Data Multivariat Dengan Program R, Jurnal Pengabdian Masyarakat MIPA dan Pendidikan MIPA*, vol. 2, no. 2, hal. 83-86.
- Yamin, Sofyan, dkk. 2011. *Regresi dan Korelasi dalam Genggaman Anda Aplikasi dengan Software SPSS, Eviews, MINITAB, dan STATGRAPHICS*. Jakarta: Salemba Empat.
- Yuwono, Margo. 2021. *Statistik Kriminal 2021*. Jakarta: Badan Pusat Statistik.