

**PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM  
*CLUSTERING* RATA-RATA PENAMBAHAN KASUS *COVID-19*  
BERDASARKAN KOTA/KABUPATEN DI PROVINSI SUMATERA  
SELATAN**

**SKRIPSI**

Sebagai salah satu syarat untuk penyelesaian studi  
di Program Studi Sistem Informasi S1



Oleh

**Sevi Dian Nirwana**

**NIM 09031181823025**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2022**

**LEMBAR PENGESAHAN**

**PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM  
*CLUSTERING* RATA-RATA PENAMBAHAN KASUS *COVID-19*  
BERDASARKAN KOTA/KABUPATEN DI PROVINSI SUMATERA  
SELATAN**

**SKRIPSI**

Sebagai salah satu syarat untuk penyelesaian studi  
di Program Studi Sistem Informasi S1

Oleh :

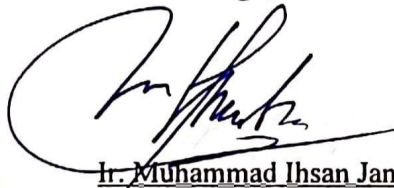
**Sevi Dian Nirwana**

**NIM 09031181823025**

Palembang, 25 Juli 2022

Pembimbing II,

Pembimbing I,



Ir. Muhammad Ihsan Jambak, M.Sc., M.M.

NIP. 196804052013081201

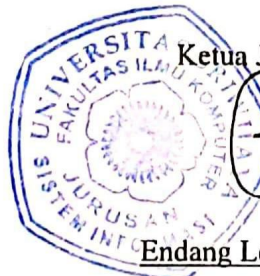


Ali Bardadi, M.Kom.

NIP. 198806292019031007

Mengetahui

Ketua Jurusan Sistem Informasi,



Endang Lestari Ruskan, S.Kom., M.T.

NIP 197811172006042001

## HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Sevi Dian Nirwana

NIM : 09031181823025

Program Studi : Sistem Informasi

Judul Skripsi : Perbandingan Algoritma *K-Means* dan *K-Medoids* dalam  
*Clustering* Rata-rata Penambahan Kasus *Covid-19*  
Berdasarkan Kota/Kabupaten di Provinsi Sumatera Selatan

Hasil Pengecekan iThenticate/Turnitin : 8 %

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri dan bukan penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Sevi Dian Nirwana

NIM.09031181823025

## HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jum'at  
Tanggal : 15 Juli 2022

Nama : Sevi Dian Nirwana  
NIM : 09031181823025  
Judul : Perbandingan Algoritma *K-Means* dan *K-Medoids* dalam  
*Clustering* Rata-rata Penambahan Kasus *Covid-19*  
Berdasarkan Kota/Kabupaten di Provinsi Sumatera Selatan

Komisi Penguji :

1. Pembimbing I : Ir. Muhammad Ihsan Jambak, M.Sc., M.M.
2. Pembimbing II : Ali Bardadi, M.Kom.
3. Ketua Penguji : Rizka Dhini Kurnia, M.Sc.
4. Penguji I : Ari Wedhasmara, M.TI.
5. Penguji II : Nabila Rizky Oktadini, M.T.



Mengetahui

Ketua Jurusan Sistem Informasi,



Endang Lestari Ruskan, S.Kom., M.T.

NIP. 197811172006042001

## HALAMAN PERSEMBAHAN

### MOTTO

***“Kerjakan Apa yang Kau Sukai dan Sukai Apa yang Sedang Kau Kerjakan”***

***“Teruslah Melangkah Dijalan yang Kau Pilih!”***  
***-River JKT48-***

***Skripsi ini aku persembahkan untuk:***

- ♥ Diriku sendiri, Sevi Dian Nirwana***
- ♥ Kedua Orang Tuaku dan 3 Saudariku Tercinta***
- ♥ Sahabat Terbaik “4Serangkai5Sempurna” dan “YOLO”***
- ♥ Teman-teman Seperjuangan SINUS A 2018***
- ♥ Dosen Pembimbing***
- ♥ Dosen dan Staff Fakultas Ilmu Komputer***
- ♥ Almamater Kebanggaanku, Universitas Sriwijaya***

## KATA PENGANTAR



Alhamdulillah rabbi'l'alamin. Segala puji dan syukur kehadiran Allah SWT atas berkah, rahmat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis, sehingga dapat menyelesaikan Skripsi ini yang berjudul **“Perbandingan Algoritma *K-Means* dan *K-Medoids* dalam *clustering* Rata-rata Penambahan Kasus *Covid-19* Berdasarkan Kota/Kabupaten di Provinsi Sumatera Selatan”**. Serta sholawat dan salam kepada Nabi Muhammad *Shallallahu ‘alaihi wasallam* sebagai sosok idola dan teladan bagi penulis.

Dalam penyusunan skripsi ini, banyak hambatan yang penulis hadapi namun pada akhirnya dapat dilalui berkat adanya bimbingan, arahan, nasihat dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu pada kesempatan ini dan dengan kerendahan hati penulis menyampaikan ucapan terimakasih kepada semua pihak, antara lain:

1. Orang tua tercinta yang telah membesarkan dan mendidik dengan penuh kasih sayang dan senantiasa memberikan doa, dukungan dan motivasi baik moral, materil maupun spiritual selama ini.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Endang Lestari Ruskan, M.T selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Nabila Rizky Oktadini, M.T selaku Pembimbing Akademik penulis yang senantiasa memberikan nasehat, motivasi dan arahan selama masa perkuliahan.

5. Bapak Ir. Muhammad Ihsan Jambak, M.Sc., M.M dan Bapak Ali Bardadi, M.Kom selaku Pembimbing Skripsi yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Skripsi ini.
6. Bapak/Ibu Dosen beserta Staff Universitas Sriwijaya khususnya Jurusan Sistem Informasi, Fakultas Ilmu Komputer.
7. Saudari penulis tercinta Anjani Putri Leontiny, Tria July Artini dan Melisa Angraini yang selalu ada, menghibur dan memberikan dukungan setiap waktu.
8. Sahabat terbaik Karisma, Deby, Melda dan Dini yang selalu menghibur dan memberikan semangat.
9. Arapa, Arin, Mao sahabat tersegalanya dari jaman maba. Terima kasih selalu kebersamai dan saling mendukung dalam perjalanan ini.
10. Teman-teman seperjuangan SINUS A 2018 dan teman-teman Organisasi di LDF WIFI, BEM Fasilkom, HIMSI dan KPU KM UNSRI yang tidak bisa penulis sebutkan satu-persatu.

Penulis menyadari bahwa Skripsi ini masih jauh dari kata kesempurnaan. Oleh karena itu kritik dan saran dari semua pihak yang bersifat membangun sangat penulis harapkan agar Skripsi ini dapat dijadikan ide dan pemikiran yang bermanfaat bagi semua pihak.

Palembang, Juli 2022  
Penulis,

Sevi Dian Nirwana  
NIM. 09031181823025

**PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM  
*CLUSTERING* RATA-RATA PENAMBAHAN KASUS *COVID-19*  
BERDASARKAN KOTA/KABUPATEN DI PROVINSI SUMATERA  
SELATAN**

Oleh

**Sevi Dian Nirwana      09031181823025**

**ABSTRAK**

Penyebaran yang cukup luas dan cepat, membuat pandemi *Covid-19* di Sumatera Selatan berdampak negatif pada semua sektor seperti kesehatan, pekerjaan dan perekonomian. Dengan kebijakan pemerintah yang mengelompokkan wilayah penanganan *Covid-19* menjadi 4 zona, perlu dievaluasi apakah pengelompokkan wilayah tersebut sudah tepat menggunakan teknik *clustering data mining* dengan algoritma *K-Means* dan *K-Medoids*. Dari hasil pengujian algoritma *K-Means* memberikan nilai DBI terbaik adalah 0.078 pada  $K=2$ . Sedangkan algoritma *K-Medoids* memberikan nilai DBI terbaik adalah 0.250 pada  $K=3$ . Sehingga kesimpulan yang didapatkan, disarankan pembagian wilayah penanganan *Covid-19* di provinsi Sumatera Selatan dibagi menjadi 2 cluster (yaitu Kota Palembang dan Luar Kota Palembang) atau menjadi 3 cluster (yaitu Kota Palembang, dekat dengan Kota Palembang dan jauh dari Kota Palembang).

**Kata Kunci :** *Covid-19, K-Means, K-Medoids, Clustering, DBI*



**COMPARISON OF K-MEANS AND K-MEDOIDS ALGORITHMS IN  
CLUSTERING AVERAGE ADDITION OF COVID-19 CASES BY  
CITY/REGENCY AT SOUTH SUMATERA PROVINCE**

By

**Sevi Dian Nirwana      09031181823025**

**ABSTRACT**

*The spread is quite wide and fast, making the Covid-19 pandemic in South Sumatra have a negative impact on all sectors such as health, employment and the economy. With the government's policy of grouping the Covid-19 handling areas into 4 zones, it is necessary to evaluate whether the grouping of these areas is appropriate using data mining clustering techniques with the K-Means and K-Medoids algorithms. From the results of testing the K-Means algorithm, the best DBI value is 0.078 at K=2. While the K-Medoids algorithm gives the best DBI value is 0.250 at K=3. So that the conclusions obtained, it is recommended that the distribution of the Covid-19 handling area in the province of South Sumatra be divided into 2 clusters (namely Palembang City and Outside Palembang City) or into 3 clusters (namely Palembang City, close to Palembang City and far from Palembang City).*

**Keywords:** Covid-19, K-Means, K-Medoids, Clustering, DBI

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>ii</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>iii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iv</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vi</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	5
1.3. Tujuan Penelitian .....	5
1.4. Manfaat Penelitian .....	6
1.5. Batasan Masalah .....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>8</b>
2.1. <i>Data Mining</i> .....	8
2.1.1. Pengertian <i>Data Mining</i> .....	8
2.1.2. Peran Utama <i>Data Mining</i> .....	9
2.2. <i>Clustering</i> .....	10
2.3. Algoritma <i>K-Means</i> .....	11
2.4. Algoritma <i>K-Medoids</i> .....	13
2.5. <i>Davies Bouldin Index (DBI)</i> .....	14
2.6. <i>Covid-19</i> .....	17
2.6.1. Pengertian <i>Covid-19</i> .....	17
2.6.2. Gejala Infeksi <i>Covid-19</i> .....	17
2.6.3. Istilah yang Berkaitan dengan <i>Covid-19</i> .....	18

2.7. <i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i> .....	20
2.8. Uji Mann-Whitney .....	22
2.9. <i>RapidMiner</i> .....	22
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>24</b>
3.1. Jenis dan Sumber Data .....	24
3.2. Tahapan Penelitian .....	24
3.3. Pengumpulan Data .....	27
3.4. <i>Data Preprocessing/Preparation</i> .....	28
3.5. Tahapan Penyelesaian Kasus .....	31
3.5.1. Tahapan Penyelesaian Kasus dengan Algoritma <i>K-Means</i> .....	31
3.5.2. Tahapan Penyelesaian Kasus dengan Algoritma <i>K-Medoids</i> .....	38
3.6. Pengolahan Data dengan <i>RapidMiner</i> .....	42
3.7. Evaluasi dan Perbandingan Kinerja Algoritma .....	44
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>46</b>
4.1. Pengujian Algoritma .....	46
4.1.1. Parameter Pengujian Algoritma .....	46
4.1.2. Hasil 13 Pengujian Algoritma dengan K=4 .....	48
4.1.3. Hasil <i>Cluster</i> K=4 .....	50
4.2. Penentuan K Optimal Masing-masing Algoritma .....	52
4.2.1. Perbandingan Hasil Pengujian pada K=2 sampai K=8 .....	52
4.2.2. <i>Centroid Table</i> dan Anggota <i>Cluster</i> .....	54
<b>BAB V PENUTUP</b> .....	<b>60</b>
5.1. Kesimpulan .....	60
5.2. Saran .....	61
<b>DAFTAR PUSTAKA</b> .....	<b>62</b>
<b>LAMPIRAN</b> .....	<b>66</b>

## DAFTAR GAMBAR

Gambar 2.1 Tahapan CRISP-DM.....	20
Gambar 3.1 Tahapan Penelitian.....	24
Gambar 3.2 Pengumpulan Data.....	28
Gambar 3.3 Variabel yang Digunakan.....	29
Gambar 3.4 Transform dan Cleaning Data.....	29
Gambar 3.5 Tahapan Algoritma <i>K-Means</i> .....	31
Gambar 3.6 Tahapan Algoritma <i>K-Medoids</i> .....	38
Gambar 3.7 <i>Import Data RapidMiner</i> .....	42
Gambar 3.8 Desain <i>Input Proses Clustering</i> .....	43
Gambar 3.9 Penentuan Jumlah <i>K</i> dan <i>Measure Types Cluster</i> .....	44
Gambar 4.1 Anggota <i>Cluster K = 4</i> .....	52
Gambar 4.2 Anggota <i>Cluster Algoritma K-Means K=2</i> .....	54
Gambar 4.3 Anggota <i>Cluster Algoritma K-Medoids K=3</i> .....	56
Gambar 4.4 Peta Cluster Penyebaran Covid-19 ( <i>K=4</i> ).....	58
Gambar 4.5 Peta <i>Cluster Penyebaran Covid-19 Algoritma K-Means K=2</i> .....	58
Gambar 4.6 Peta <i>Cluster Penyebaran Covid-19 Algoritma K-Medoids K=3</i> .....	59

## DAFTAR TABEL

Tabel 3.1 Data Final Rata-rata Penambahan Kasus.....	30
Tabel 3.2 Hasil Iterasi 1 Algoritma <i>K-Means</i> .....	32
Tabel 3.3 Hasil Iterasi 2 Algoritma <i>K-Means</i> .....	33
Tabel 3.4 Hasil Iterasi 3 Algoritma <i>K-Means</i> .....	34
Tabel 3.5 Hasil Iterasi 4 Algoritma <i>K-Means</i> .....	35
Tabel 3.6 Hasil Iterasi 5 Algoritma <i>K-Means</i> .....	36
Tabel 3.7 Hasil Iterasi 1 Algoritma <i>K-Medoids</i> .....	39
Tabel 3.8 Hasil Iterasi 2 Algoritma <i>K-Medoids</i> .....	40
Tabel 4.1 Jenis Parameter Pengujian.....	46
Tabel 4.2 Hasil 13 Pengujian Algoritma dengan $K=4$ .....	47
Tabel 4.3 Statistik Deskriptif DBI dan <i>Avg Within Cluster Distance</i> .....	47
Tabel 4.4 Hasil Uji Normalitas DBI.....	48
Tabel 4.5 Rank DBI Uji Mann-Whitney.....	48
Tabel 4.6 <i>Test Statistics</i> Uji Mann-Whitney.....	49
Tabel 4.7 <i>Test of Homogeneity of Variance Avg Within Cluster Distance</i> .....	49
Tabel 4.8 <i>Centroid K = 4 dengan Measure Types Numerical Measure, Divergence ChebychevDistance</i> .....	50
Tabel 4.9 Perbandingan <i>K Cluster Algoritma (MT = Numerical Measure, Divergence=ChebychevDistance)</i> .....	53
Tabel 4.10 <i>Centroid K-Means K=2</i> .....	54
Tabel 4.11 <i>Centroid K-Medoids K=3</i> .....	55

## DAFTAR LAMPIRAN

Lampiran 1 Hasil Pengujian dengan SPSS.....	A-1
Lampiran 2 Persentase Hasil Turnitin.....	B-1
Lampiran 3 Kartu Konsultasi Skripsi.....	C-1
Lampiran 4 Surat Keputusan (SK) TA.....	D-1
Lampiran 5 Form Perbaikan Ujian Komprehensif.....	E-1

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Indonesia dan hampir semua negara di dunia saat ini masih terus berjuang melawan pandemi virus yang belum bisa dipastikan kapan akan berakhir. *World Health Organization* (WHO) memberi nama virus tersebut *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2) atau Virus Corona dengan nama penyakitnya adalah *Coronavirus Disease 2019* atau lebih familiar dengan sebutan *Covid-19* (Yuliana, 2020). Pertama mewabah di China di akhir tahun 2019 tepatnya di bulan Desember. Virus menular dengan cepat dan menyebabkan infeksi saluran pernapasan ringan hingga dapat menyebabkan orang yang terpapar meninggal dunia. WHO menyatakan pandemi *Covid-19* pada tanggal 11 Maret 2020. (Indraputra & Fitriana, 2020). Awal tahun 2020 merupakan tahun dimana pandemi *Covid-19* tepatnya pada tanggal 2 Maret 2020, mulai masuk ke Indonesia dengan kasus pertamanya pasien yang terinfeksi sebanyak dua orang dan terus bertambah setiap harinya.(Dwitri dkk, 2020). Kasus *Covid-19* telah menyebar merata ke seluruh provinsi di Indonesia. Salah satu provinsi yang tak luput dari pandemi *Covid-19* adalah Sumatera Selatan. Pandemi ini telah menyebar ke berbagai Kota dan Kabupaten se-Sumatera Selatan. Terdapat 17 wilayah kabupaten/kota yang menjadi titik penyebaran virus ini yaitu Kabupaten Banyuasin, Kabupaten Empat Lawang, Kabupaten Lahat, Kabupaten Muara Enim, Kabupaten Musi Banyuasin, Kabupaten Musi Rawas, Kabupaten Musi Rawas Utara, Kabupaten Ogan Ilir, Kabupaten

Ogan Komering Ilir, Kabupaten Ogan Komering Ulu, Kabupaten Ogan Komering Ulu Selatan, Kabupaten Ogan Komering Ulu Timur, Kabupaten Pematang Abab Lematang Ilir, Kota Lubuklinggau, Kota Pagar Alam, Kota Palembang dan Kota Prabumulih (Dinkes Sumsel, 2020).

Penyebaran yang cukup luas dan cepat, membuat pandemi *Covid-19* berdampak negatif pada semua sektor. Tidak hanya menimbulkan korban jiwa, dampak krisis kesehatan, pekerjaan dan perekonomian juga dirasakan Indonesia termasuk Sumatera Selatan. Dengan didasari pada kebijakan pemerintah yang membagi wilayah penanganan *Covid-19* menjadi 4 zona wilayah yaitu hijau, kuning, oranye dan merah, dalam hal ini ingin melihat dan mengevaluasi apakah pembagian wilayah tersebut sudah tepat dan sesuai dengan realita data *Covid-19* yang ada. Sehingga disini perlunya mengelompokkan wilayah persebaran *Covid-19* di Sumatera Selatan agar bisa disesuaikan dengan kondisi *Covid-19* pada saat ini yang divalidasi dari kondisi data *Covid-19* yang sebenarnya.

Dalam hal ini penulis memanfaatkan salah satu metodologi dalam ilmu komputer yaitu *Data Mining* untuk menganalisis dan menggali informasi dari data yang ada. *Data Mining* memiliki beberapa teknik analisis data seperti *clustering*, *classification*, *prediction*, dan *association*. *Clustering* merupakan teknik *data mining unsupervised learning* yang berperan untuk mengelompokkan sekumpulan *object* data kedalam beberapa *cluster* tanpa terikat dengan label yang mengaturnya. *Object* yang mempunyai kemiripan akan dikumpulkan menjadi *cluster* yang sama, dan *object* dengan karakteris-



tik berbeda akan dipisahkan ke *cluster* yang berbeda. (Hariyanto & Shita, 2018).

Penelitian sebelumnya yang dilakukan oleh Sindi, dkk (2020) menjelaskan bahwa melalui teknik *clustering* dapat diketahui pola pemilihan penentuan pengelompokan penyebaran *Covid-19* di berbagai wilayah di Indonesia. Disamping itu, penelitian oleh Mirantika, dkk (2021) melakukan *clustering* kasus *Covid-19* di Provinsi Jawa Barat menggunakan algoritma *K-Means* yang mendapatkan hasil *clustering* sebanyak 3 *cluster* penyebaran *Covid-19* dengan menggunakan variabel jumlah isolasi, sembuh dan meninggal dunia. Penelitian dengan kasus serupa juga dilakukan oleh Bu'ulolo dan Purba (2021) dengan algoritma yang berbeda untuk membentuk *cluster* zona penyebaran *Covid-19* di Provinsi Sumatera Utara dengan algoritma *K-Medoids* menggunakan variabel positif, ODP, PDP dan meninggal. Hasil yang didapatkan dari penelitian ini yaitu didapatkannya 3 *cluster* zona penyebaran.

Dari beberapa penelitian diatas telah memberikan hasil yg cukup baik dalam *clustering* penyebaran kasus *Covid-19* terhadap wilayah yang menjadi subjek penelitian tersebut. Akan tetapi penelitian serupa yang memanfaatkan *data mining* belum dilakukan di Provinsi Sumatera Selatan yang juga merupakan daerah terdampak *Covid-19*. Selain itu, dengan dugaan bahwa kasus *Covid-19* memiliki penyebaran dan penambahan kasus yang tidak merata dan memiliki selisih penyebaran yang sangat jauh antara wilayah centrum penyebaran *Covid-19* dengan wilayah diluar centrumnya dengan begitu sangat mungkin jika data *Covid-19* bersifat anomali. Maka dalam hal

ini, digunakan dua algoritma yang berbeda untuk membandingkan serta mencari validasi terhadap *cluster* yang terbentuk sebagai evaluasi lebih lanjut terhadap algoritma yang digunakan tersebut benar-benar optimal dan cukup ideal dalam proses *clustering* sesuai dengan kondisi data.

Pada penelitian ini, penulis menguji dan membandingkan dua algoritma dari metode *Partitional clustering* yaitu *K-Means* dan *K-Medoids* dalam mengelompokkan rata-rata penambahan kasus *Covid-19* di Provinsi Sumatera Selatan dengan data yang sama. Algoritma *K-Means* yang digunakan karena kinerjanya yang relatif cepat, mudah diimplementasikan dan mudah beradaptasi akan dibandingkan dengan Algoritma *K-Medoids* yang dinilai lebih unggul dan optimal pada data yang berjumlah sedikit serta bersifat *noise* dan *outlier* yaitu ketika objek bernilai besar menyimpang dari distribusi data. (Wahyudi dkk, 2020). Persamaan dari dua algoritma ini yaitu harus menentukan jumlah *k* atau jumlah *cluster* terlebih dahulu lalu kemudian menentukan titik *centroid* masing-masing *cluster*. Algoritma *K-Means* menentukan nilai *centroid* menggunakan nilai rata-rata, sedangkan Algoritma *K-Medoids* menggunakan nilai tengah atau *median* (Sangga, 2018). Dengan perbedaan, keunggulan dan kelemahan masing-masing algoritma ini membuat pengujian algoritma dilakukan untuk melihat perbandingan apakah terdapat perbedaan yang signifikan dari hasil *clustering* dua algoritma tersebut meskipun keduanya masih berkaitan.

Berdasarkan penjelasan diatas, maka penulis tertarik mengambil judul penelitian yaitu **“Perbandingan Algoritma *K-Means* dan *K-Medoids*”**

**dalam *clustering* Rata-rata Penambahan Kasus *Covid-19* Berdasarkan Kota/Kabupaten di Provinsi Sumatera Selatan”.**

## **1.2. Rumusan Masalah**

Rumusan masalah penelitian ini yaitu algoritma apa yang lebih optimal diantara *K-Means* dan *K-Medoids* dalam *clustering* rata-rata penambahan kasus *Covid-19* di Provinsi Sumatera Selatan. Untuk menjawab pertanyaan pada rumusan masalah tersebut, maka diuraikan menjadi beberapa pertanyaan yaitu :

1. Bagaimana tahapan dari algoritma *K-Means* dan algoritma *K-Medoids* dalam melakukan *clustering* ?
2. Bagaimana perbandingan algoritma *K-Means* dan algoritma *K-Medoids* dalam *clustering* rata-rata penambahan kasus *Covid-19* di Provinsi Sumatera Selatan ?
3. Berapa jumlah *cluster* terbaik dari masing-masing algoritma dan karakteristik atau variabel apa yang menyebabkan pengelompokan dalam *clustering* rata-rata penambahan kasus *Covid-19* di Provinsi Sumatera Selatan ?

## **1.3. Tujuan Penelitian**

Tujuan yang hendak dicapai penulis dalam penelitian ini adalah

1. Untuk mengetahui dan memahami tahapan dan perhitungan menggunakan algoritma *K-Means* dan algoritma *K-Medoids* saat melakukan *clustering*.
2. Untuk membandingkan dan mengetahui algoritma yang lebih optimal antara algoritma *K-Means* dan *K-Medoids* dalam *clustering* rata-rata penambahan kasus *Covid-19* di Provinsi Sumatera Selatan.

3. Untuk mengetahui karakteristik atau variabel yang menjadi penyebab pengelompokan data kedalam suatu *cluster*.

#### **1.4. Manfaat Penelitian**

Manfaat yang didapatkan dari penelitian ini dibagi menjadi manfaat praktis dan akademis. Adapun manfaat yang dimaksud adalah sebagai berikut :

##### **1. Manfaat Praktis**

Melalui penelitian ini diharapkan dapat menjadi evaluasi ataupun rekomendasi kepada pemerintah sebagai pertimbangan dalam mengelompokkan wilayah penyebaran kasus *Covid-19* di provinsi Sumatera Selatan agar penanganannya dapat lebih efektif.

##### **2. Manfaat Akademis**

Menambah wawasan baru bagi civitas akademika khususnya terhadap penerapan *data mining*, serta sebagai referensi yang dapat digunakan pada penelitian *data mining* yang akan datang. Terutama dalam melakukan penelitian yang relevan menggunakan teknik *clustering*.

#### **1.5. Batasan Masalah**

Untuk menghindari agar pembahasan dalam penelitian ini tidak menyimpang, maka penulis membatasi permasalahan sebagai berikut :

1. Penelitian ini akan menggunakan data penambahan kasus *Covid-19* di Sumatera Selatan selama 6 bulan terhitung sejak bulan April hingga September 2021, diambil dari *website* resmi Dinas Kesehatan Provinsi Sumatera Selatan.

2. Variabel data yang akan digunakan dalam *clustering* ini adalah variabel nama kota/kabupaten, suspek, kontak erat, sembuh, meninggal dan dirawat.
3. Penelitian ini menggunakan teknik *clustering data mining* untuk menguji dua algoritma yaitu *K-Means* dan *K-Medoids*.
4. Tahapan pada penelitian ini mengadopsi metode CRISP-DM.
5. Hasil *clustering* yang dibandingkan dari kedua algoritma ini yaitu nilai *Davies Bouldin Index* (DBI) dan nilai *Average within cluster Distance*.
6. *Tools* yang digunakan untuk pengolahan data adalah *RapidMiner*.

## DAFTAR PUSTAKA

- Aperdanaste, J., Hasanah, A., & Siregar, A. Y. M. (2022). Pengaruh Pembatasan Sosial dan Variabel Ekonomi Terhadap Prevalensi Pandemi Covid-19 di Provinsi Jawa Barat. *Jurnal Ekonomi Kesehatan Indonesia*, 6(2). <https://doi.org/10.7454/eki.v6i2.5218>
- Bates, A. & Kalita, J. (2016). Counting clusters in twitter posts. *Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies*, pp. 85.
- Bu'ulolo, E., & Purba, B. (2021). Algoritma *clustering* Untuk Membentuk *cluster* Zona Penyebaran Covid-19. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 12(1), 59–67. <https://doi.org/10.31849/digitalzone.v12i1.6572>
- Darmansah, D. D., & Wardani, N. W. (2021). Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode *K-Means clustering*. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 105–117. <https://doi.org/10.35957/jatisi.v8i1.590>
- Dwitri, N., dkk. (2020). Penerapan Algoritma *K-Means* Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia. *Jti (Jurnal Teknologi Informasi)*, 4(1), 128–132.
- Elvitaria, L. & Havenda, M. (2017). Memprediksi Tingkat Peminat Ekstrakurikuler pada Siswa Smk Analisis Kesehatan Abdurrah Menggunakan Algoritma C4.5 (Studi Kasus: SMK Analisis Kesehatan Abdurrah). *RABIT(Jurnal Teknologi dan Sistem Informasi Univrab)*, 2(2), 220–233.
- Farahdinna, F., dkk. (2019). Perbandingan Algoritma *K-Means* Dan *K-Medoids* Dalam Klasterisasi Produk Asuransi Perusahaan Nasional. *Jurnal Ilmiah FIFO*, 11(2), 208. <https://doi.org/10.22441/fifo.2019.v11i2.010>
- Hariyanto, M., & Shita, R. T. (2018). *clustering* pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma *K-Means* dan Metode Perhitungan Jarak Euclidean Distance. *Sistem Komputer Dan Teknik Informatika*, 1(1), 117–122.
- Indraputra, R. A., & Fitriana, R. (2020). *K-Means clustering* Data COVID-19. *Jurnal Teknik Industri*.10(3), 275–282.
- Jambak, M. I., & Efendi, R. (2021). Pengaruh Reduksi Dimensi Terhadap Metode Pengklasteran Berbasis Centroid dan Metode Pengklasteran Berbasis Density dalam Pengklasteran Dokumen Teks. *Indonesian Journal of Business Intelligence*, 4(2), 53–62. <http://dx.doi.org/10.21927/ijubi.v4i2.1918>

- Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma *K-Means*. *Jurteks*, 4(2), 123–132. <https://doi.org/10.33330/jurteks.v4i2.36>
- Marlina, D., dkk. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Mirantika, Nita, A. T. diviana agnia. (2021). *Penerapan Algoritma K-Means clustering Untuk Pengelompokan Penyebaran Covid-19 Jurnal Nuansa Informatika Volume 15 Nomor 2, Juli 2021. 15*, 92–98.
- Nawrin, S., Rahman, M.R. & Akhter, S. (2017). Exploreing k-means with internal validity indexes for data clustering in traffic management system. *International Journal of Advanced Computer Science and Applications*, 8(3): 264-272.
- Novalia, Dhea. (2021). *Pengujian Algoritma K-Means Dalam Klasterisasi Data Persediaan Obat di Klinik Sahabat Mandiri*. Universitas Sriwijaya.
- Nugraha, D. D. C., dkk. (2014). Klasterisasi Judul Buku dengan Menggunakan Metode K-Means. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 21(1), 1907–5022.
- Quraisy, A., & Madya, S. (2021). Analisis Nonparametrik Mann Whitney Terhadap Perbedaan Kemampuan Pemecahan Masalah Menggunakan Model Pembelajaran Problem Based Learning. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 3(1), 51–57. <https://doi.org/10.35580/variansiunm23810>
- Rahmat, B., dkk (2017). Implementasi K-Means Clustering pada RapidMiner untuk Analisis Daerah Rawan Kecelakaan. *Seminar Nasional Riset Kuantitatif Terapan*, 58-63.
- Ramageri, Bharati M. (2010). Data Mining Techniques and Applications. *Indian Journal of Computer Science and Engineering*, 1(4), 301–305.
- Sangga, V. A. P. (2018). *Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah Tahun 2015*. Tugas Akhir Jurusan Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Inndonesia Yogyakarta.
- Sari, R. W., dkk (2018). Aplikasi *RapidMiner* dalam Pengelompokan Kasus Penyakit AIDS berdasarkan Provinsi dengan Data Mining *K-Means clustering*. *Regional Development Industry & Health Science, Technology and Art of Life*, 59–69.

<https://ptki.ac.id/jurnal/index.php/readystar/article/viewFile/7/pdf> (05 Juni 2020)

- Sari, Y. R., dkk. (2020). Penerapan Algoritma *K-Means* Untuk *clustering* Data Kemiskinan Provinsi Banten Menggunakan *RapidMiner*. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 192. <https://doi.org/10.24114/cess.v5i2.18519>
- Sholihah, S. A. (2021). *Analisis cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means*. Universitas Islam Negeri Sunan Ampel.
- Sihombing, E. G. (2017). Klasifikasi Data Mining pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan *K-Means clustering Method*. *Computer Engineering, System and Science Journal*, 2(2), 74–82.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Ilmi R.H.Zer, F., & Hartama, D. (2020). Analisis algoritma *K-Medoids clustering* dalam pengelompokan penyebaran Covid-19 di Indonesia. *Jti (Jurnal Teknologi Informasi)*, 4(1), 166–173.
- Sitompul, Bernad, J.D. (2018). *Tesis:peningkatan hasil evaluasi clustering davies bouldin dengan penentuan titik pusat cluster awal K means*. Universitas Sumatera Utara.
- Solichin, A., & Khairunnisa, K. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode *K-Means*. *Fountain of Informatics Journal*, 5(2), 52. <https://doi.org/10.21111/fij.v5i2.4905>
- Susilo, A., dkk (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), 45. <https://doi.org/10.7454/jpdi.v7i1.415>
- Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode *K-Means* Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 11(2), 119–128. <https://doi.org/10.30598/barekengvol11iss2pp119-128>
- Utomo, W. (2021). The comparison of *K-Means* and *K-Medoids* algorithms for *clustering* the spread of the covid-19 outbreak in Indonesia. *ILKOM Jurnal Ilmiah*, 13(1), 31–35. <https://doi.org/10.33096/ilkom.v13i1.763.31-35>
- Wani, M. A. & Riyaz, R. (2017). A novel point density based validity index for clustering gene expression datasets. *International Journal of Data Mining and Bioinformatics* 17(1): 66–84.



Waworuntu, N. V., & Amin, M. F. (2018). Penerapan Metode *K-Means* Pemetaan Calon. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 05(02), 190–200.

Yuliana. (2020). Wellness and Healthy Magazine. *Journal Wellnes*, 2(1), 187–192.  
<https://wellnes.journalpress.id/wellnes>