

**ANALISIS KLASTER UNTUK DATA KATEGORIK MENGGUNAKAN
METODE *K-MODES*
(Studi Kasus: Data Pasien COVID-19 di RSUD Dr. H. Abdul Moeloek
Provinsi Lampung)**

(Skripsi)

Oleh

SHABRINA NOVAINDAH DWIYAMTI



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

**ANALYSIS CLUSTER FOR CATEGORICAL DATA USING K-MODES
METHOD
(STUDY CASE: COVID-19 PATIENTS DATA AT RSUD DR. H. ABDUL
MOELOEK LAMPUNG PROVINCE)**

By

SHABRINA NOVAINDAH DWIYAMTI

Cluster analysis is a multivariate technique with interdependence because there is no difference between independent variable and dependent variable. Cluster analysis consists of hierarchical and non hierarchical methods. K-Means is a non hierarchical cluster analysis method. However, the K-Means method is limited to numerical data. So we need a method to process the categorical data. One of the non hierarchical methods for categorical data often used is K-Modes. This study aims to apply K-Modes cluster analysis to COVID-19 patients data at RSUD Dr. H. Abdul Moeloek Lampung Province, totaling 560 patient data with variables of gender, age, admission, and their condition when leaving RSUD Dr. H. Abdul Moeloek Lampung Province. By using Davies-Bouldin Index (DBI) and Silhouette method, the optimal k value for the K-Modes cluster analysis is 8 clusters, consisting of 145 patients in the first cluster, 227 patients in the second cluster, 16 patients in the third cluster, 30 patients in the fourth cluster, 30 patients in the fifth cluster, 74 patients in the sixth cluster, 4 patients in the seventh cluster, dan 34 patients in the eighth cluster. Because there are more members of clusters 1 and 2 when compared to other clusters, more optimal handling is needed for clusters 1 and 2.

Keywords: Cluster Analysis, K-Modes, COVID-19, DBI, Silhouette Method.

**ANALISIS KLASTER UNTUK DATA KATEGORIK MENGGUNAKAN
METODE *K-MODES*
(STUDI KASUS: DATA PASIEN COVID-19 DI RSUD DR. H. ABDUL
MOELOEK PROVINSI LAMPUNG)**

Oleh

SHABRINA NOVAINDAH DWIYAMTI

Analisis klaster merupakan salah satu analisis multivariat metode interdependensi dikarenakan tidak ada perbedaan antara variabel bebas dan variabel tak bebas. Analisis klaster terdiri dari metode hierarki dan non hierarki. *K-Means* merupakan salah satu metode analisis klaster non hierarki. Namun, metode *K-Means* terbatas pada data numerik. Sehingga dibutuhkan metode untuk mengolah data kategorik. Salah satu metode non hierarki untuk data kategorik yang sering digunakan adalah *K-Modes*. Penelitian ini bertujuan untuk menerapkan analisis klaster *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung yang berjumlah 560 data pasien dengan variabel jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Dengan menggunakan Davies-Bouldin *Index* (DBI) dan metode *Silhouette*, diperoleh hasil nilai k optimal untuk analisis klaster *K-Modes* adalah sebanyak 8 klaster, yaitu terdiri dari 145 pasien klaster 1, 227 pasien klaster 2, 16 pasien klaster 3, 30 pasien klaster 4, 30 pasien klaster 5, 74 pasien klaster 6, 4 pasien klaster 7, dan 34 pasien klaster 8. Karena anggota klaster 1 dan 2 terbanyak jika dibandingkan dengan klaster lainnya, maka diperlukan penanganan yang lebih optimal untuk klaster 1 dan 2.

Kata Kunci: Analisis Klaster, *K-Modes*, COVID-19, DBI, Metode *Silhouette*.

**ANALISIS KLASTER UNTUK DATA KATEGORIK MENGGUNAKAN
METODE *K-MODES*
(Studi Kasus: Data Pasien COVID-19 di RSUD Dr. H. Abdul Moeloek
Provinsi Lampung)**

Oleh

Shabrina Novaindah Dwiyantri

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

Judul Skripsi : **ANALISIS KLASTER UNTUK DATA KATEGORIK MENGGUNAKAN METODE K-MODES (Studi Kasus: Data Pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung)**

Nama Mahasiswa : **Shabrina Novaindah Dwiyamti**

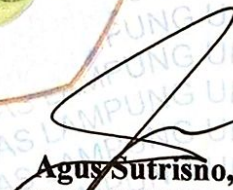
Nomor Pokok Mahasiswa : **1817031009**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

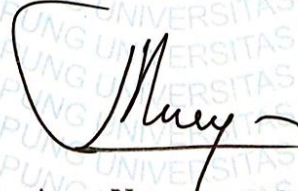


Dr. Khoirin Nisa, S.Si., M.Si.
NIP 197407262000032001



Agus Sutrisno, S.Si., M.Si.
NIP 197008311999031002

2. Ketua Jurusan Matematika



Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.**



Sekretaris : **Agus Sutrisno, S.Si., M.Si.**



Penguji
Bukan Pembimbing : **Ir. Netti Herawati, M.Sc., Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Satripto Dwi Yuwono, S.Si., M.T.

NIP. 19740705 200003 1 001



Tanggal Lulus Ujian Skripsi : 10 Juni 2022

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama mahasiswa : **SHABRINA NOVAINDAH DWIYAMTI**
Nomor Pokok Mahasiswa : **1817031009**
Jurusan : **Matematika**
Judul Skripsi : **Analisis Klaster untuk Data Kategorik**
Menggunakan Metode K-Modes
(Studi Kasus: Data Pasien COVID-19 di RSUD
Dr. H. Abdul Moeloek Provinsi Lampung)

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, Juni 2022

Yang menyatakan,



Shabrina Novaindah Dwiyamti

RIWAYAT HIDUP

Penulis bernama lengkap Shabrina Novaindah Dwiyamti. Lahir di Bandar Lampung pada tanggal 06 November 2000, merupakan anak kedua dari tiga bersaudara, pasangan Bapak Nyamin dan Ibu Muhayati. Penulis mempunyai kakak bernama Luthfi Dermawan Bagustiyan dan adik bernama Fadhil Rafi Rahmatullah.

Penulis mengawali Pendidikan kanak-kanak di TK Al-Huda Bandar Lampung pada tahun 2005-2006. Kemudian penulis melanjutkan pendidikan Sekolah Dasar di SD Negeri 1 Langkapura pada tahun 2006-2012. Selanjutnya penulis melanjutkan Pendidikan Sekolah Menengah Pertama di SMP Negeri 14 Bandar Lampung pada tahun 2012-2015 dan melanjutkan pendidikan Sekolah Menengah Atas di SMA Negeri 9 Bandar Lampung pada tahun 2015-2018.

Pada tahun 2018 penulis melanjutkan pendidikan Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung. Selama menjadi mahasiswa penulis aktif di beberapa organisasi yaitu Generasi Muda HIMATIKA (GEMATIKA) 2018 dan Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila sebagai Anggota Bidang Eksternal periode 2020.

Pada bulan Februari hingga Maret 2021 penulis melaksanakan Kuliah Kerja Nyata (KKN) Periode I di Kelurahan Langkapura, Kecamatan Langkapura, Kota Bandar Lampung sebagai bentuk pengabdian kepada masyarakat. Selanjutnya pada bulan Juli hingga Agustus 2021 penulis melaksanakan Kerja Praktik (KP) di Badan Pendapatan Daerah (BAPENDA) Provinsi Lampung sebagai bentuk pengembangan diri serta menerapkan ilmu yang telah didapat selama perkuliahan.

KATA INSPIRASI

*“Maka sesungguhnya bersama kesulitan ada kemudahan,
sesungguhnya bersama kesulitan ada kemudahan”*
(QS. Al-Insyirah: 5-6)

*“Allah tidak membebani seseorang melainkan sesuai dengan
kesanggupannya”*
(QS. Al-Baqarah: 286)

*“Dia yang pergi untuk mencari ilmu pengetahuan, dianggap
sedang berjuang di jalan Allah sampai dia kembali”*
(HR. Tirmidzi)

*“You do not just wake up and become the butterfly, growth is
a process”*
(Rupi Kaur)

“Everything is possible. The impossible just takes longer”
(Dan Brown)

PERSEMBAHAN

Puji dan syukur saya haturkan kepada Allah SWT. Yang telah memberikan rahmat, hidayah, dan karunia-Nya kepada saya. Saya persembahkan karya sederhana ini untuk:

Ibu, Bapak, Abang, dan Adik

Terima kasih telah memberikan doa dan dukungan yang tiada henti kepada penulis.

Dosen Pembimbing dan Dosen Pembahas

Terima kasih telah meluangkan waktu untuk memberikan bimbingan, kritik, saran, dan ilmu sehingga skripsi ini dapat terselesaikan.

Sahabat-sahabatku

Terima kasih telah memberikan doa dan dukungan kepada penulis, baik dalam proses perkuliahan maupun dalam proses penulisan skripsi.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji dan syukur penulis haturkan kepada Allah SWT., yang telah memberikan rahmat, hidayah, serta karunia-Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Klaster untuk Data Kategorik Menggunakan Metode *K-Modes* (Studi Kasus: Data Pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung)”.

Terselesaikannya skripsi ini tidak lepas dari dukungan, bimbingan, saran, serta doa dari berbagai pihak. Dengan segala kerendahan hati penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Khoirin Nisa, S.Si, M.Si., selaku Dosen Pembimbing 1 atas kesabaran dan kesediaannya untuk memberikan bimbingan, kritik, dan saran dalam proses penyelesaian skripsi ini serta selalu meluangkan waktunya untuk bimbingan.
2. Bapak Agus Sutrisno, S.Si., M.Si., selaku Dosen Pembimbing 2 yang telah memberikan saran serta arahan kepada penulis dan selalu meluangkan waktunya untuk bimbingan.
3. Ibu Ir. Netti Herawati, M.Sc. P.hD., selaku Dosen Pembahas skripsi yang telah memberikan kritik, saran dan masukan yang sangat membantu penulis dalam memperbaiki skripsi ini.
4. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Dosen Pembimbing Akademik yang telah memberikan arahan dan bantuannya dalam masa perkuliahan sehingga penulis dapat menyelesaikan perkuliahan dengan baik.
5. Bapak Dr. Aang Nuryaman, S.Si, M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

6. Bapak Dr. Eng Suropto Dwi Yuwono, M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen, staf, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah banyak membantu selama perkuliahan.
8. Seluruh staf bagian Pendidikan Pengembangan SDM & Hukum RSUD Dr. H. Abdul Moeloek Provinsi Lampung yang telah memberikan kesempatan untuk pengambilan data dan mendukung penelitian skripsi ini.
9. Ibu, Bapak, Abang Luthfi Dermawan Bagustiyani, dan Adik Fadhil Rafi Rahmatullah yang selalu memberi doa dan dukungan selama perkuliahan serta penulisan skripsi ini.
10. Sahabat-sahabatku sejak mahasiswa baru, yaitu Faricha Salsabila Azzahra, Shofiyah Fauziah Sayuti, Nadya Aristiawati Sitorus, Nanda Ane Famelia Atika, Lutfia Humairoso, Mutia Lutfi Gunawan, dan Syava Zahwa yang selalu menemani sejak mahasiswa baru dan selama pengerjaan skripsi ini.
11. Maydia Egi Nuraini, Virda Kurniawati, dan Maulana Yusuf yang selalu memberikan dukungan serta memberi saran atas penelitian skripsi ini.
12. Teman-teman satu bimbingan, yaitu Sofalina Nodra Brilliantya, Syifaa Aalimatul Haqqi, dan Rexi Soaloon Pakpahan yang selalu memberikan semangat dalam pengerjaan skripsi ini.
13. Teman-teman “Matematika A” terima kasih atas kebersamaannya.
14. Teman-teman Matematika 2018, terima kasih atas kebersamaannya.
15. Seluruh pihak yang telah membantu dan terlibat dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa banyak terdapat kekurangan dalam penulisan skripsi ini masih. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak

Bandar Lampung, Juni 2022
Penulis

Shabrina Novaindah Dwiyamti

DAFTAR ISI

	Halaman
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah.....	1
1.2 Tujuan Penelitian	2
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	4
2.1 Analisis Multivariat	4
2.2 Analisis Klaster	4
2.3 Metode Hierarki	5
2.4 Metode Non Hierarki	7
2.4.1 Metode <i>K-Modes</i>	8
2.5 Validasi Klaster.....	10
2.5.1 Davies-Bouldin <i>Index</i> (DBI).....	10
2.5.2 Metode <i>Silhouette</i>	12
III. METODOLOGI PENELITIAN	13
3.1 Waktu dan Tempat Penelitian.....	13
3.2 Data Penelitian	13
3.3 Metode Penelitian	14
IV. HASIL DAN PEMBAHASAN	16
4.1 Statistika Deskriptif Data Penelitian	16
4.2 Analisis Klaster Menggunakan Metode <i>K-Modes</i>	17
4.3 Hasil Pengklasteran.....	19
4.4 Penentuan Jumlah Klaster Terbaik	20
4.4.1 Davies-Bouldin <i>Index</i> (DBI).....	20
4.4.2 Metode <i>Silhouette</i>	23
4.5 Interpretasi dan Profiling Hasil Klaster untuk $k = 8$	24

V. KESIMPULAN.....	27
DAFTAR PUSTAKA	28
LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Variabel yang Digunakan pada Data Penelitian.....	14
2. Objek yang Dipilih sebagai <i>Modes</i> Awal.....	18
3. Hasil Pengklasteran $k = 2, 3, 4, 5, 6, 7, 8$	19
4. Nilai DBI untuk $k = 2, 3, 4, 5, 6, 7, 8$	23
5. Karakteristik Setiap Klaster untuk $k = 8$	24

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Pie Chart</i> Data Penelitian.....	16
2. Grafik Nilai DBI	23
3. Visualisasi k Optimal Metode <i>Silhouette</i>	24

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Statistika merupakan ilmu pengetahuan yang berkaitan dengan data dan fakta yang tepat. Statistika meliputi teknik pengumpulan data, pengolahan, penganalisisan, penarikan kesimpulan, serta pembuatan keputusan yang tepat berdasarkan fakta yang ada (Wanitaningsih, 2012).

Dalam mengelompokkan objek-objek berdasarkan kemiripan variabel-variabel yang diamati dapat pula digunakan analisis statistika. Salah satu analisis untuk menyelesaikan masalah tersebut, yaitu analisis kluster. Menurut Santosa dkk. (2019), analisis kluster adalah salah satu analisis multivariat yang dapat digunakan untuk meringkas dan mempermudah interpretasi data. Analisis kluster bertujuan untuk mengelompokkan objek-objek ke dalam kelompok-kelompok yang relatif homogen (Wibowo dkk., 2020).

Menurut Johnson & Wichern (2007), analisis kluster terdiri dari metode hierarki dan non hierarki. Metode hierarki digunakan jika belum ada informasi jumlah kluster yang akan dipilih. Sedangkan, metode non hierarki jumlah kluster sudah ditentukan sebelumnya.

K-Means adalah salah satu metode analisis kluster non hierarki. Pengklasteran *K-Means* sangat sesuai jika digunakan pada data dengan ukuran yang besar dikarenakan mempunyai perhitungan yang cepat (Ananta, 2017). Namun, metode *K-Means* terbatas pada data numerik. Sehingga dibutuhkan metode untuk

mengolah data kategorik. Salah satu metode non hierarki untuk data kategorik yang sering digunakan adalah *K-Modes*.

Metode *K-Modes* mula-mula diperkenalkan oleh Huang pada tahun 1997. Metode *K-Modes* adalah modifikasi dari metode *K-Means* (Aranganayagi & Thangavel, 2009). Metode *K-Modes* termasuk ke dalam metode partisi yang membagi objek ke-*k* kelompok yang berbeda sehingga objek yang memiliki karakteristik yang sama masuk ke dalam satu klaster yang sama dan objek yang memiliki karakteristik berbeda masuk ke dalam klaster yang lain.

Metode *K-Modes* dapat menghasilkan klaster melalui proses yang lebih detail, waktu yang dibutuhkan untuk komputasi pembentukan klaster lebih cepat, dan sangat baik digunakan dalam klasterisasi pada data berdimensi banyak (Kartikasari dkk., 2021).

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian mengenai analisis klaster menggunakan metode *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung, penulis menggunakan Davies-Bouldin *Index* (DBI) dan metode *Silhouette* untuk mengevaluasi jumlah klaster optimal.

1.2 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah menerapkan analisis klaster *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung.

1.3 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Dapat menambah wawasan tentang penerapan ilmu matematika dalam kehidupan sehari-hari.
2. Dapat memperoleh hasil kluster terbaik untuk pengklasteran pasien COVID-19 berdasarkan jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung.

II. TINJAUAN PUSTAKA

2.1 Analisis Multivariat

Analisis statistika yang digunakan pada data yang terdiri atas beberapa variabel dan antarvariabel saling berkorelasi disebut analisis multivariat. Data multivariat mencakup tidak hanya satu variabel saja, tetapi terdiri dari beberapa variabel (Alwi & Hasrul, 2018).

Menurut Sitepu dkk. (2011), analisis multivariat terbagi menjadi dua jenis, yaitu:

1. Metode dependensi. Jika tujuan dari analisis adalah menguraikan atau memperkirakan variabel tak bebas berdasarkan dua ataupun lebih variabel bebas, maka digunakan metode dependensi. Metode ini terdiri atas empat jenis, yaitu analisis diskriminan berganda, analisis multivariat varians, dan analisis korelasi kanonikal.
2. Metode interdependensi. Jika tujuan dari analisis adalah menjelaskan beberapa variabel atau pengelompokan berdasarkan beberapa variabel tertentu, maka digunakan metode interdependensi. Metode ini terdiri dari tiga jenis, yaitu analisis faktor, analisis kluster, dan skala multidimensional.

2.2 Analisis Kluster

Metode mengelompokkan kumpulan objek ke dalam satu atau lebih kelompok sehingga data yang tergabung pada suatu kelompok memiliki tingkat kesamaan disebut analisis kluster (Rohmatullah dkk., 2019).

Analisis kluster adalah teknik analisis yang berfungsi untuk mengelompokkan objek-objek berdasarkan ukuran kemiripan ataupun ketidakmiripan. Semakin mirip dua objek, maka akan semakin tinggi kemungkinan berada dalam suatu kluster. Demikian juga, semakin berbeda karakteristik suatu objek, maka kemungkinan untuk berada dalam suatu kluster semakin kecil (Chrisinta dkk., 2020).

Menurut Muhibbuddin dkk. (2013), ciri-ciri kluster yang baik adalah sebagai berikut:

1. Homogenitas (kesamaan) yang besar antarobjek dalam satu kluster.
2. Heterogenitas (perbedaan) yang besar antarkluster yang satu dengan kluster yang lainnya.

2.3 Metode Hierarki

Terdapat dua teknik dalam analisis kluster hierarki, yaitu penggabungan (*agglomerative*) dan pemecahan (*divisive*). Menurut Goreti dkk. (2016), analisis kluster menggunakan metode hierarki diawali dengan mengelompokkan objek yang memiliki kemiripan yang paling dekat. Lalu dilanjutkan ke objek lain yang memiliki kedekatan kedua. Demikian selanjutnya sehingga kelompok akan terbentuk seperti “pohon”. Akan ada tingkatan yang detail antarobjek, dari yang paling mirip sampai dengan yang paling tidak mirip (Pramudita & Sumargo, 2019)

Menurut Rachmatin (2014), terdapat beberapa metode analisis kluster hierarki, yaitu:

1. Metode pautan tunggal (*single linkage method*)

Metode pautan tunggal merupakan metode pengklasteran berdasarkan pada jarak terdekat antarobjek. Apabila dua objek terpisah oleh jarak yang dekat, maka kedua objek tersebut akan berkelompok menjadi satu kluster dan demikian selanjutnya. Metode pautan tunggal dilakukan dengan cara mengasumsikan setiap objek dianggap sebagai kluster, selanjutnya menentukan

klaster yang memiliki jarak terdekat, misalnya klaster A dan klaster B memiliki jarak terdekat lalu dikelompokkan, hasil pengelompokannya disebut dengan klaster AB . Dari klaster AB yang sudah terbentuk hitung jarak minimum antarklaster AB dengan objek lainnya yang belum dikelompokkan, misalnya objek C . Metode pautan tunggal dapat dicari menggunakan rumus sebagai berikut:

$$d_{(AB)C} = \min (d_{AC}, d_{BC})$$

dengan:

d_{AC} = jarak antara klaster A dan C

d_{BC} = jarak antara klaster B dan C

2. Metode pautan penuh (*complete linkage method*)

Metode pautan penuh merupakan metode pengklasteran berdasarkan pada jarak terjauh antarobjek. Apabila dua objek terpisah oleh jarak yang jauh, maka kedua objek tersebut akan bergabung menjadi satu klaster dan demikian selanjutnya. Metode pautan penuh dilakukan dengan cara mengasumsikan setiap objek dianggap sebagai klaster, selanjutnya menentukan klaster yang memiliki jarak terjauh, misalnya klaster A dan klaster B memiliki jarak terjauh lalu dikelompokkan, hasil pengelompokannya disebut dengan klaster AB . Dari klaster AB yang telah terbentuk hitung jarak maksimum antarklaster AB dengan objek lainnya yang belum dikelompokkan, misalnya objek C . Metode pautan penuh dapat dicari menggunakan rumus sebagai berikut:

$$d_{(AB)C} = \max (d_{AC}, d_{BC})$$

dengan:

d_{AC} = jarak antara klaster A dan C

d_{BC} = jarak antara klaster B dan C

3. Metode pautan rata-rata (*average linkage method*)

Metode pautan rata-rata merupakan metode pengklasteran berdasarkan pada jarak rata-rata antarobjek. Metode pautan rata-rata dilakukan dengan cara mengasumsikan setiap objek dianggap sebagai klaster, selanjutnya menentukan klaster yang memiliki jarak terdekat, misalnya klaster A dan klaster B memiliki jarak terdekat lalu dikelompokkan, hasil pengelompokannya disebut dengan klaster AB . Dari klaster AB yang telah terbentuk hitung jarak rata-rata

antarklaster AB dengan objek lainnya yang belum dikelompokkan, misalnya objek C . Metode pautan rata-rata dapat dicari menggunakan rumus sebagai berikut:

$$d_{(AB)C} = \frac{d_{AC} + d_{BC}}{n_{AB} n_C}$$

dengan:

d_{AC} = jarak antara klaster A dan C

d_{BC} = jarak antara klaster B dan C

n_{AB} = banyaknya anggota klaster AB

n_C = banyaknya anggota klaster C

4. Metode Ward

Dalam metode Ward, jarak antara dua klaster diperoleh dari total jumlah kuadrat dua klaster pada setiap variabel. Metode Ward berbeda dengan metode yang lain karena memakai pendekatan analisis varians untuk menghitung jarak antarklaster atau metode ini meminimumkan jumlah kuadrat.

5. Metode *centroid linkage*

Dalam metode *centroid linkage*, jarak antara dua klaster diperoleh dari jarak antar *centroid* klaster tersebut. *Centroid* klaster merupakan nilai tengah observasi pada variabel dalam suatu set variabel klaster. Keuntungan metode *centroid linkage* adalah *outlier* tidak banyak berpengaruh apabila dibandingkan dengan metode lainnya.

2.4 Metode Non Hierarki

Analisis klaster metode non hierarki memiliki kecepatan yang lebih tinggi dibandingkan dengan metode hierarki dan lebih menguntungkan jika jumlah sampel besar. Tetapi terdapat kekurangan dari metode non hierarki, yaitu jumlah klaster sudah ditetapkan sebelumnya dan pusat klaster dipilih secara sembarang. Hasil klaster akan tergantung pada bagaimana pusat dipilih. Sehingga, hasil klaster mungkin bergantung pada urutan observasi dalam data (Mahmudah & Yordani, 2016).

Menurut Musfiani (2019), jenis-jenis metode dalam analisis kluster non hierarki adalah *sequential threshold*, *parallel threshold*, dan *optimizing partitioning* (metode partisi optimasi).

Metode partisi dilakukan dengan menentukan jumlah awal kluster, kemudian menghitung ulang jarak setiap objek. Apabila pada perhitungan ulang jarak objek lebih mendekati kluster lain dibandingkan kluster saat ini, maka harus dilakukan penempatan kembali objek-objek ke dalam kluster yang lebih dekat. Jadi, dalam metode partisi sebuah objek dapat dimasukkan ke dalam sebuah kluster, kemudian dipindahkan karena ternyata lebih dekat dengan kluster yang lain (Wilandari dkk., 2014).

Saat mempartisi satu set objek menjadi k kluster, tujuan utamanya adalah untuk mendapatkan kluster berdasarkan objek yang memperlihatkan tingkat kesamaan yang tinggi (Musfiani, 2019).

2.4.1 Metode *K-Modes*

Dalam analisis kluster non hierarki terdapat metode yang terkenal karena efisiensinya dalam mengelompokkan kumpulan data yang besar, yaitu metode *K-Means*. Namun, karena keterbatasan ukuran ketidakmiripan yang dimiliki oleh metode *K-Means*, maka metode *K-Means* tidak dapat digunakan pada data kategorik (nominal atau ordinal). Persoalan tersebut bisa diatasi oleh metode *K-Modes* yang merupakan modifikasi dari metode *K-Means*. Metode *K-Modes* adalah metode yang paling dikenal dalam pengelompokan kumpulan data kategorik berjumlah besar ke kluster tertentu yang diwakili oleh modus atau nilai yang paling sering muncul (Khan & Ahmad, 2013).

Menurut Az-zahra dkk. (2021), tahapan yang dapat dilakukan dalam menjalankan analisis kluster metode *K-Modes* adalah sebagai berikut:

1. Menentukan banyaknya kluster (k) yang akan dibentuk dari r data. Jumlah data harus lebih banyak dibandingkan dengan jumlah kluster ($r > k$).
2. Menentukan k modus awal dari data yang akan digunakan untuk pusat kluster (*centroid*). Banyaknya *centroid* yang dipilih harus sesuai dengan banyaknya kluster yang akan dibentuk. Contohnya, apabila ingin membentuk dua kluster, maka banyaknya *centroid* adalah dua.
3. Mencari jarak setiap objek terhadap setiap *centroid*.

Jarak dicari dengan memakai ukuran ketidakmiripan sederhana (*simple matching dissimilarity*). Asumsikan $B1$ dan $B2$ merupakan dua data yang mempunyai n variabel kategorik. Ukuran ketidakmiripan antara $B1$ dan $B2$ dapat didefinisikan menjadi total ketidakcocokan dari variabel-variabel kategorik yang sesuai dari kedua data. Semakin mirip kedua data, ditandai dengan nilai ketidakcocokan yang semakin kecil. Persamaan ukuran ketidakmiripan sederhana dirumuskan sebagai berikut:

$$d(V_1, V_2) = \sum_{j=1}^n \delta(x_{1j}, x_{2j}) \quad (2.3)$$

dengan:

$$\delta(x_{1j}, x_{2j}) = \begin{cases} 0, & x_{1j} = x_{2j} \\ 1, & x_{1j} \neq x_{2j} \end{cases}$$

x_{1j} = nilai dari variabel ke- j pada objek V_1

x_{2j} = nilai dari variabel ke- j pada objek V_2

4. Menempatkan objek berdasarkan jarak terdekat ke *centroid* menggunakan rumus jarak terdekat. Kluster mana yang paling dekat dengan objek dapat diketahui dengan nilai ukuran ketidakmiripan sederhana.
5. Apabila seluruh data sudah ditempatkan ke dalam kluster terdekat, kemudian memperbarui *centroid* setiap kluster berdasarkan modus dari setiap variabel anggota kluster yang terbentuk.
6. Mencari kembali jarak setiap data terhadap se *centroid* baru menggunakan ukuran ketidakmiripan sederhana seperti tahapan (3). Apabila pada penghitungan ulang jarak, data lebih mendekati kluster lain dibandingkan

klaster saat ini, maka data ditempatkan kembali ke klaster lain yang lebih dekat tersebut.

7. Mengulangi tahapan (5) dan (6) sampai tidak ada data yang berpindah klaster.

2.5 Validasi Klaster

2.5.1 Davies-Bouldin Index (DBI)

Metode Davies-Bouldin Index (DBI) mula-mula diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Menurut Wani & Riyaz (2017), salah satu metode yang berfungsi untuk mengukur validitas klaster pada suatu metode pengklasteran adalah DBI. Pengukuran dengan DBI ini memaksimalkan jarak inter-klaster antara klaster C_k dan C_l dan meminimalkan jarak antartitik dalam sebuah klaster. Apabila jarak inter-klaster maksimal, artinya kesamaan karakteristik antar setiap klaster sedikit sehingga perbedaan antar-klaster terlihat lebih rinci. Apabila jarak intra-klaster minimal artinya setiap objek dalam klaster tersebut mempunyai tingkat kesamaan karakteristik yang tinggi.

Menurut Az-zahra dkk. (2021), rumus menghitung DBI adalah sebagai berikut:

$$DBI = \frac{1}{p} \times \sum_{c=1}^p R_c \quad (2.4)$$

dengan:

$$R_c = \max_{c \neq d} R_{cd} \quad (2.5)$$

dan

$$R_{cd} = \frac{S_c + S_d}{d(V_c, V_d)} \quad (2.6)$$

dengan:

DBI = Davies-Bouldin Index

p = banyaknya klaster

R_{cd} = ukuran kemiripan antara klaster ke- c dan klaster ke- d

S_c = ukuran dispersi pada klaster ke- c

S_d = ukuran dispersi pada klaster ke- d

$$S_c = \left[\frac{1}{n_c} \sum_{T_k \in C_c, k=1}^{n_c} d^2(T_k, V_c) \right]^{\frac{1}{2}} \quad (2.7)$$

$$d^2(T_k, V_c) = (d(T_k, V_c))^2$$

dengan:

n_c = banyaknya anggota klaster ke- c , $c = 1, 2, \dots, p$

C_c = klaster ke- c

T_k = anggota ke- k pada klaster ke- c

V_c = *centroid* klaster ke- c

$d(T_k, V_c)$ = jarak dari anggota ke- k pada klaster ke- c (T_k) dengan *centroid* klaster ke- c (V_c)

Nilai $d(T_k, V_c)$ dapat dihitung memakai ukuran ketidakmiripan pencocokan sederhana seperti pada persamaan berikut:

$$d(T_k, V_c) = \sum_{l=1}^m \delta(x_{kl}, v_{cl}) \quad (2.8)$$

dengan:

$$\delta(x_{kl}, v_{cl}) = \begin{cases} 0, & x_{kl} = v_{cl} \\ 1, & x_{kl} \neq v_{cl} \end{cases}$$

dengan:

x_{kl} = nilai dari variabel ke- l pada T ke- k

v_{cl} = nilai ke- l pada *centroid* variabel ke- c

m = banyaknya variabel

Rumus untuk mendapatkan nilai varians klaster ke- d juga sama dengan rumus untuk menghitung jarak *centroid* klaster ke- c (V_c) ke *centroid* klaster ke- d (V_d) sebagai berikut:

$$d(V_c, V_d) = \sum_{l=1}^n \delta(v_{cl}, v_{dl}) \quad (2.9)$$

dengan:

$$\delta(v_{cl}, v_{dl}) = \begin{cases} 0, & v_{cl} = v_{dl} \\ 1, & v_{cl} \neq v_{dl} \end{cases}$$

dengan:

v_{dl} = nilai ke- l pada *centroid* variabel ke- d

2.5.2 Metode *Silhouette*

Metode *Silhouette* adalah metode yang berfungsi untuk mengetahui kualitas dan kekuatan dari klaster. Metode *Silhouette* adalah campuran dari dua metode yaitu metode *cohesion* yang dapat digunakan dalam menghitung seberapa dekat relasi antara objek dalam sebuah klaster, dan metode *separation* yang dapat digunakan dalam mengukur seberapa jauh sebuah klaster terpisah dengan klaster lain (Pramesti dkk., 2017). Langkah-langkah untuk mendapatkan nilai *Silhouette* adalah sebagai berikut:

1. Pada setiap objek k cari rata-rata jarak dari objek k dengan semua objek yang berada dalam klaster yang sama, yaitu $c(k)$ dan cari rata-rata jarak dari objek k dengan objek yang berada pada klaster yang lainnya, yaitu $d(k)$. Lalu cari nilai *Silhouette* objek, yaitu $sil(x_i)$

$$sil(x_i) = \frac{d(k) - c(k)}{\max \{c(k), d(k)\}}$$

2. Untuk setiap klaster ke- l , cari rata-rata dari klaster objek *Silhouette* dengan menggunakan persamaan

$$sil(C_j) = \frac{1}{|C_l|} \sum_{x_k \in C_l} sil(x_i)$$

3. Hitung nilai *Silhouette* dengan n adalah banyaknya klaster

$$sil(k) = \frac{1}{|n|} \sum_{r=1}^n sil(C_l)$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2021/2022 dan bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder mengenai pasien COVID-19. Data tersebut diperoleh dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Variabel yang digunakan pada penelitian ini, yaitu jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung.

Pada penelitian ini digunakan 560 data pasien, yaitu 461 data pasien pada bulan Agustus 2021, 69 data pasien pada bulan September 2021, dan 30 data pasien pada bulan Oktober 2021. Data yang digunakan adalah data kategorik. Data pada penelitian ini harus diubah ke dalam bentuk skala pengukuran nominal terlebih dahulu. Kemudian dibentuk klaster sebanyak 2, 3, 4, 5, 6, 7, dan 8. Selanjutnya dihitung nilai validitas dari hasil pengklasteran untuk melihat keefektifan dari metode *K-Modes*.

Tabel 1. Variabel yang Digunakan pada Data Penelitian

Variabel		Kategorisasi
X_1	Jenis Kelamin	0; Laki-Laki 1; Perempuan
X_2	Usia	1; 0-19 tahun 2; 20-39 tahun 3; 40-59 tahun 4; 60-79 tahun 5; 80-99 tahun
X_3	Cara Masuk	0; Rujukan dokter RS 1; Datang sendiri
X_4	Kondisi Saat Keluar dari RS	1; Sembuh 2; Belum sembuh 3; Berobat jalan 4; Meninggal < 48 jam 5; Meninggal \geq 48 jam

3.3 Metode Penelitian

Data pada penelitian ini diolah menggunakan aplikasi *R-Studio* dengan tahapan sebagai berikut:

1. Menentukan nilai k , pada penelitian ini digunakan $k = 2, 3, 4, 5, 6, 7, 8$
2. Menentukan *centroid* (pusat kluster) dengan cara acak dari setiap kluster.
3. Mencari jarak tiap objek terhadap *centroid* menggunakan ukuran ketidakmiripan sederhana.
4. Menempatkan objek berdasarkan jarak terdekat ke *centroid*.
5. Mencari modus dari setiap variabel sebagai *centroid* kluster.
6. Menentukan ulang jarak tiap objek terhadap *centroid* baru menggunakan ukuran ketidakmiripan sederhana.

7. Menempatkan objek berdasarkan jarak terdekat ke *centroid*.
8. Jika terdapat objek yang berpindah kluster, maka ulangi tahapan ke lima sampai tujuh hingga tidak ada objek yang berpindah kluster.
9. Menentukan k terbaik menggunakan DBI dan metode *Silhouette*.
10. Menentukan profiling dari setiap kluster.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan pada Bab IV, maka dapat diambil kesimpulan sebagai berikut:

1. Analisis kluster metode *K-Modes* merupakan metode yang dapat digunakan untuk tipe data kategorik.
2. Pengelompokan data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung terbagi menjadi 8 kluster berdasarkan jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Dari hasil penelitian, dominasi kluster 1 dan kluster 2 cukup signifikan karena setiap beranggotakan 145 dan 227 dari 560 pasien.

DAFTAR PUSTAKA

- Alwi, W. & Hasrul, M. 2018. Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Propinsi Sulawesi Selatan berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Matematika dan Statistika serta Aplikasinya*. **6**(1): 35-42.
- Ananta, A. Y. 2017. Seleksi Notifikasi Serangan Berbasis IDS Snort Menggunakan Metode K-Means. *SMARTICS Journal*. **3**(2): 31-38.
- Aranganayagi, S. & Thangavel, K. 2009. Improved K-Modes for Categorical Clustering Using Weighted Dissimilarity Measure. *International Journal of Computer and Information Engineering*. **3**(3): 729-735.
- Az-zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila., & Madjida, W. O. Z. 2021. Penerapan Algoritma K-Modes Klastering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Matematika dan Statistika serta Aplikasinya*. **9**(1): 24:36.
- Chrisinta, D., Sumertajaya, I. M., & Indahwati. 2020. Evaluasi Kinerja Metode Klaster Ensemble dan Latent Class Klastering pada Peubah Campuran. *Indonesian Journal of Statistics and Its Applications*. **4**(3): 448-461.
- Goreti, M., Nasution, Y. N., & Wahyuningsih, S. 2016. Perbandingan Hasil Analisis Klaster dengan Menggunakan Metode Single Linkage dan Metode C-Means (Studi Kasus: Data Tingkat Kualitas Udara Ambien pada Perusahaan Perkebunan di Kabupaten Kutai Barat Tahun 2014). *Jurnal Eksponensial*. **7**(1): 9-16.
- Johnson, R.A. & Wichern, D. 2007. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.

- Kartikasari, Y. A. P., Pranoto, Y. A., & Rudhistiar, D. 2021. Penerapan Metode K-Modes untuk Proses Penentuan Penerima Bantuan Langsung Tunai (BLT). *Jurnal Mahasiswa Teknik Informatika*. **5**(1): 389-397.
- Khan, S. S. & Ahmad, A. 2013. Cluster Center Initialization Algorithm For K-Modes Clustering. *Expert Systems with Applications*. **40**(18): 7444-7456.
- Mahmudah, H. & Yordani, R. 2016. Visualisasi Penggerombolan Wilayah Berdasarkan Teori Pertumbuhan Ekonomi Menggunakan Aplikasi Integrasi Self Organizing Map (SOM) dan Sistem Informasi Geografis. *Jurnal Aplikasi Statistika & Komputasi Statistik*. **7**(2): 129-142.
- Muhibbuddin, Irawan M. I., & Subchan. 2013. Pendekatan Fuzzy Clustering untuk Mengetahui Potensi Usaha mikro di Surabaya. *Jurnal Sains dan Seni Pomits*. **2**(1): 1-7.
- Musfiani. 2019. Analisis Cluster dengan Menggunakan Metode Partisi pada Pengguna Alat Kontrasepsi di Kalimantan Barat. *Buletin Ilmiah Math Stat dan Terapannya*. **8**(4): 893-902.
- Pramesti, D. F., Furqon, M. T., & Dewi, C. 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*. **1**(9): 723-732.
- Pramudita, D. A. & Sumargo, B. 2019. Pengelompokan Pengguna Internet dengan Metode K-Means Clustering. *Jurnal Statistika dan Aplikasinya*. **3**(1): 1-12.
- Rachmatin, D. 2014. Aplikasi Metode-metode Agglomerative Dalam Analisis Klaster pada Data Tingkat Polusi Udara. *Infinity Journal*. **3**(2): 133-149.
- Rohmatullah, A., Rahmalia, D., & Pradana, M. S. 2019. Klasterisasi Data Pertanian di Kabupaten Lamongan Menggunakan Algoritma K-Means dan Fuzzy C-Means. *Jurnal Ilmiah Teknosains*. **5**(2): 86-93.

- Santosa, S., Panjidang, W. L., & Santosa, Y. P. 2019. Klasterisasi Kecerdasan Majemuk Siswa Berbasis Jaringan Syaraf Kohonen Guna Mendukung Adaptive Elearning. *Jurnal Teknologi Informasi*. **15**(2): 123-141.
- Sitepu, R., Irmeilyana, & Gultom, B. 2011. Analisis Klaster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. **14**(3): 11-17.
- Wanitaningsih, S. K. 2012. Program Statistik: Upaya Memotivasi Mahasiswa untuk Senang Belajar Statistika. *Media Bina Ilmiah*. **6**(4): 62-67.
- Wani, M. A. & Riyaz, R. 2017. A Novel Point Density Based Validity Index for Clustering Gene Expression Datasets. *Int. J. Data Mining and Bioinformatics*. **17**(1): 66-84.
- Wibowo, R. A., Nisa, K., Faisol, A., & Setiawan, E. 2020. Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage dan Ward pada Data yang Mengandung Masalah Multikolinearitas. *Jurnal Siger Matematika*. **1**(2): 49-55.
- Wilandari, Y., Mukid, M. A., Megawati, N., & Sutarno, Y. A. 2014. Analisis Klasifikasi Kabupaten Di Jawa Tengah Berdasarkan Populasi Ternak Menggunakan Fuzzy Cluster Means. *Media Statistika*. **7**(2): 77-88.