

**ANALISIS ALGORITMA K-MEANS DALAM PENGELOMPOKKAN
PERSEBARAN COVID-19
DI INDONESIA**

SKRIPSI



disusun oleh

Nurul Khasanah Fitriyani

19.21.1333

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS ALGORITMA K-MEANS DALAM PENGELOMPOKKAN
PERSEBARAN COVID-19
DI INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Nurul Khasanah Fitriyani

19.21.1333

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS ALGORITMA K-MEANS DALAM PENGELOMPOKAN
PERSEBARAN COVID-19
DI INDONESIA**

yang dipersiapkan dan disusun oleh

Nurul Khasanah Fitriyani

19.21.1333

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Juli 2021

Dosen Pembimbing,

Ferian Fauzi Abdulloh, M.Kom

NIK. 190302276

PENGESAHAN
SKRIPSI
ANALISIS ALGORITMA K-MEANS DALAM PENGELOMPOKAN
PERSEBARAN COVID-19
DI INDONESIA

yang dipersiapkan dan disusun oleh

Nurul Khasanah Fitriyani

19.21.1333

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Juli 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, M.Kom

NIK. 190302185

Rizqi Sukma Kharisma, M.Kom

NIK. 190302215

Ferian Fauzi Abdulloh, M.Kom

NIK. 190302276

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 15 Agustus 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 15 Agustus 2021



Nurul Khasanah Fitriyani

NIM. 19.21.3333

MOTTO

Barang siapa yang menghendaki kehidupan dunia maka wajib baginya memiliki ilmu, dan barang siapa yang menghendaki kehidupan akherat, maka wajib baginya memiliki ilmu, dan barang siapa menghendaki keduanya maka wajib memiliki ilmu.

(HR. Turmudzi)

Kita dinilai bukan dari apa yang kita mulai. Melainkan dari apa yang kita selesaikan.

(Bong Chandra)

Selesaikan apa yang sudah kamu mulai. Berhenti ditengah jalan, bukanlah jiwa seorang pemenang.

Boleh Ngeluh tapi Jangan Nyerah!

(Nurul Khasanah Fitriyani)

PERSEMBAHAN

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan Skripsi ini. Saya ingin menyampaikan terimakasih yang setulus-tulusnya kepada :

1. Ibu saya Sagiye dan ayah saya Ngadiyono yang telah mendoakan saya dimanapun saya berada dan selalu mendukung saya sehingga saya dapat menyelesaikan Tugas Akhir ini.
2. Kakak saya Samilah dan Yuni Lestari yang selalu mendukung saya serta semua keluarga dan saudara-saudara saya yang telah mendoakan saya.
3. Dosen pembimbing saya, Bapak Ferian Fauzi Abdulloh, M.Kom yang telah membimbing dan mengarahkan saya dalam menyelesaikan Skripsi ini.
4. Dosen di Universitas AMIKOM Yogyakarta, terimakasih untuk bimbingan dan ilmu yang telah diberikan kepada saya.
5. Teman-teman saya 16-D3TI-03 dan S1-Informatika Transfer 2019 yang telah memberikan dukungan dan doanya.
6. Sahabat saya Aprida, Nupus, Diyan, Dalillah, Agnes dan teman-teman kerja saya, terimakasih sudah membantu dukungan semangat sehingga skripsi ini bisa terselesaikan.
7. Dan semua pihak yang tidak dapat disebutkan satu-persatu, saya ucapkan terima kasih.

KATA PENGANTAR

Segala puji dan syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan Skripsi ini dengan judul “**Analisis Algoritma K-Means dalam Pengelompokan Penyebaran Covid-19 di Indonesia**”.

Skripsi ini disusun sebagai syarat kelulusan Program Studi Sarjana Komputer di Universitas AMIKOM Yogyakarta.

Selama penyusunan Skripsi ini, penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, Skripsi ini tidak mungkin dapat terselesaikan. Oleh karena itu, penulis menyampaikan terima kasih kepada:

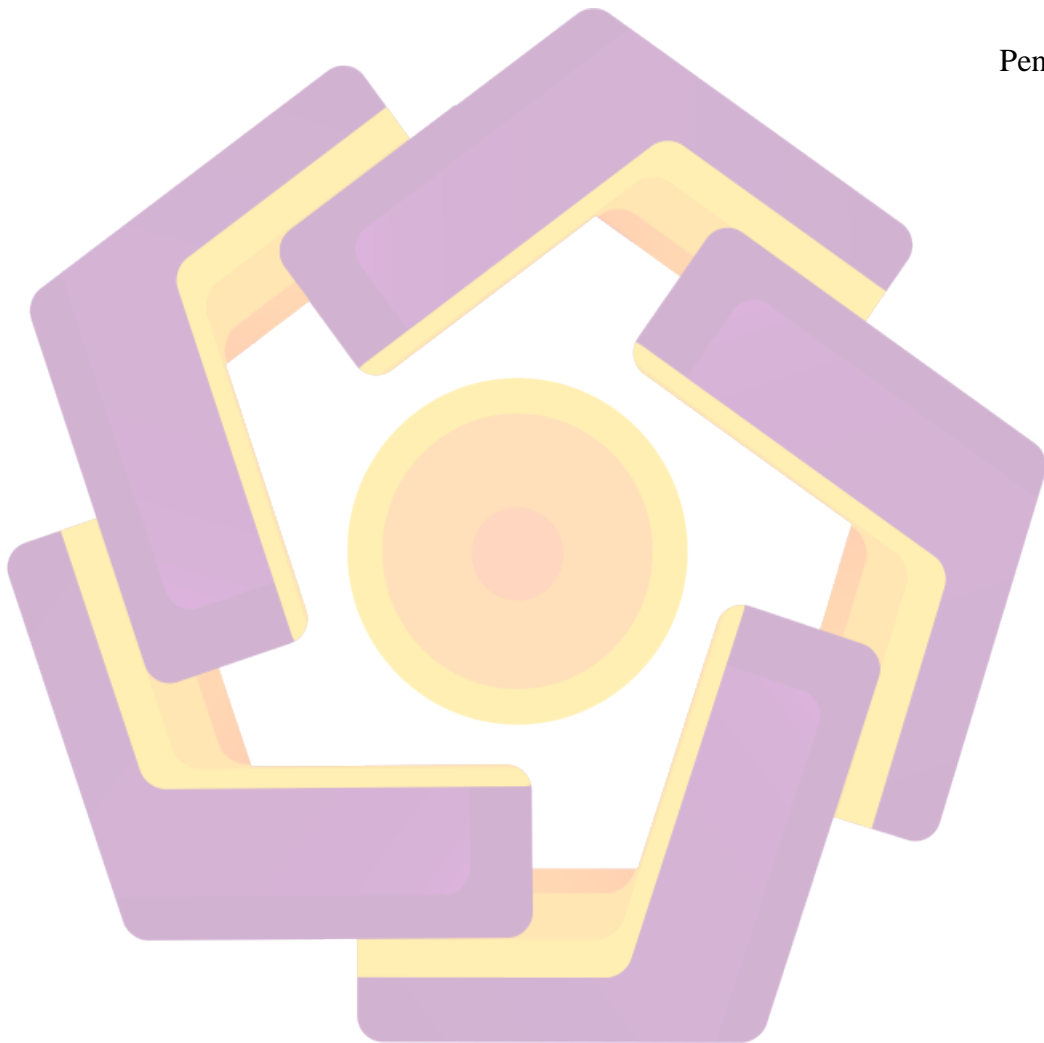
1. Allah SWT yang telah memberikan terang-Nya kepada penulis sehingga dapat menyelesaikan Tugas Akhir ini.
2. Bapak Prof. Dr. M. Suyanto, M.M selaku rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
4. Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Program Studi S1 Informatika dan Dosen Wali saya.
5. Bapak Ferian Fauzi Abdulloh, M.Kom. selaku dosen pembimbing.
6. Seluruh Dosen, Staff, maupun Karyawan Universitas AMIKOM Yogyakarta yang telah memberikan ilmu, dukungan dan sarana.
7. Kedua orang tua penulis yang selalu mendoakan dan memberikan dukungan baik dari segi moril maupun materi bagi penulis.
8. Keluarga S1-Informatika Transfer 2019 dan teman-teman di Universitas AMIKOM Yogyakarta.
9. Semua pihak yang telah membantu selama proses pembuatan Skripsi ini.

Penulis sadar bahwa masih terdapat banyak kesalahan dan kekurangan. Untuk itu penulis memohon maaf yang sebesar-besarnya atas hal tersebut. Akhir

kata, semoga skripsi ini bisa bermanfaat bagi para pembacanya, institusi pendidikan, dan khususnya bagi penulis sendiri.

Yogyakarta, 15 Agustus 2021

Penulis



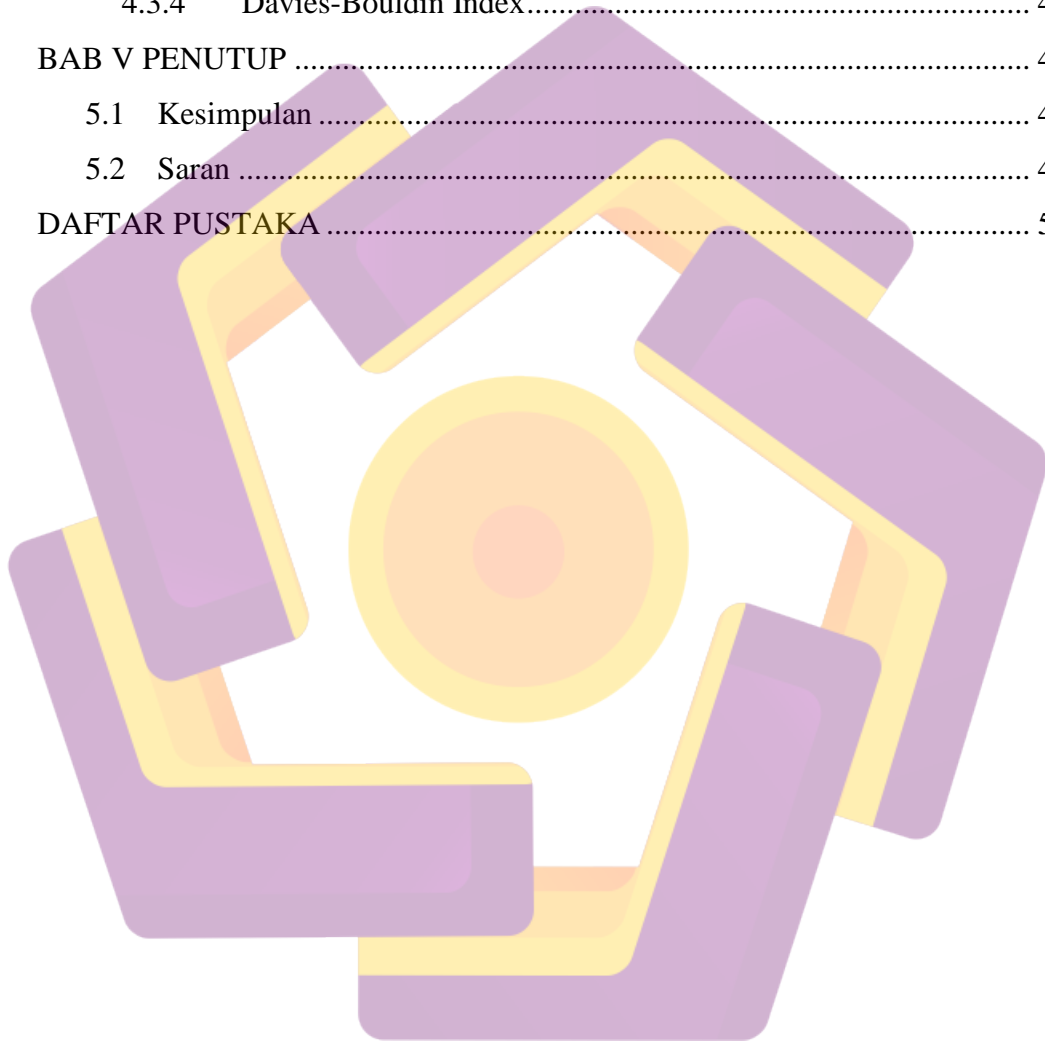
DAFTAR ISI

| | |
|--|------|
| JUDUL | i |
| PERSETUJUAN | ii |
| PENGESAHAN | iii |
| PERNYATAAN | iv |
| MOTTO | v |
| PERSEMBAHAN | vi |
| KATA PENGANTAR | vii |
| DAFTAR ISI | ix |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| INTISARI | xv |
| ABSTRACT | xvi |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Batasan Masalah | 2 |
| 1.4 Maksud dan Tujuan Penelitian | 2 |
| 1.5 Manfaat Penelitian | 3 |
| 1.6 Metode Penelitian | 3 |
| 1.7 Sistematika Penulisan | 4 |
| BAB II LANDASAN TEORI | 5 |
| 2.1 Tinjauan Pustaka | 5 |
| 2.2 Data Mining | 6 |
| 2.2.1 Data Selection | 7 |
| 2.2.2 Pre-processing | 7 |
| 2.2.3 Transformation | 7 |
| 2.2.4 Data Mining | 7 |
| 2.2.5 Interpretation/Evaluation | 7 |
| 2.3 Pengelompokan Data Mining | 8 |

| | | |
|--|---|-----------|
| 2.3.1 | Clustering..... | 8 |
| 2.3.2 | Classification..... | 8 |
| 2.3.3 | Association..... | 8 |
| 2.3.4 | Estimations..... | 8 |
| 2.3.5 | Predictions..... | 9 |
| 2.4 | Algoritma K-Means Clustering..... | 9 |
| 2.5 | Konsep Dasar Sistem..... | 10 |
| 2.5.1 | Pengertian Sistem..... | 10 |
| 2.5.2 | Karakteristik Sistem..... | 10 |
| 2.6 | Entity Relationship Diagram (ERD)..... | 11 |
| 2.7 | Data Flow Diagram (DFD)..... | 12 |
| 2.8 | Microsoft Excel..... | 13 |
| 2.9 | MySQL..... | 13 |
| BAB III ANALISIS DAN PERANCANGAN..... | | 14 |
| 3.1 | Deskripsi..... | 14 |
| 3.2 | Metode Penelitian..... | 14 |
| 3.3 | Analisis Kebutuhan..... | 14 |
| 3.3.1 | Analisis Kebutuhan Perangkat Lunak..... | 14 |
| 3.3.2 | Analisis Kebutuhan Perangkat Keras..... | 15 |
| 3.4 | Tahapan Penelitian..... | 15 |
| 3.4.1 | Pengumpulan Data..... | 15 |
| 3.4.2 | Preprocessing Data..... | 15 |
| 3.4.3 | K-Means Clustering..... | 15 |
| 3.5 | Perancangan Sistem..... | 18 |
| 3.5.1 | Perancangan DFD..... | 18 |
| 3.5.1.1 | DFD Level 0..... | 18 |
| 3.5.1.2 | DFD Level 1..... | 19 |
| 3.5.2 | Perancangan ERD..... | 20 |
| 3.5.3 | Perancangan Tabel Database..... | 21 |
| 3.5.3.1 | Tabel Admin..... | 21 |
| 3.5.3.2 | Tabel Centroid..... | 21 |

| | | |
|--|---|-----------|
| 3.5.3.3 | Tabel Objek | 22 |
| 3.5.3.4 | Tabel Satuan | 22 |
| 3.5.4 | Perancangan Interface | 22 |
| 3.5.4.1 | Rancangan Halaman Login..... | 23 |
| 3.5.4.2 | Rancangan Halaman Home..... | 24 |
| 3.5.4.3 | Rancangan Halaman Data..... | 25 |
| 3.5.4.4 | Rancangan Halaman Proses Clustering | 26 |
| BAB IV HASIL DAN PEMBAHASAN | | 27 |
| 4.1 | Penerapan K-Means Clustering menggunakan Microsoft Excel | 27 |
| 4.1.1 | Pengumpulan Data..... | 27 |
| 4.1.2 | Preprocessing..... | 27 |
| 4.1.3 | Penerapan Algoritma K-Means Clustering menggunakan Microsoft Excel 2010 | 28 |
| 4.1.3.1 | Menentukan Titik Pusat Cluster..... | 28 |
| 4.1.3.2 | Menghitung Jarak Data ke Setiap Cluster..... | 28 |
| 4.1.3.3 | Mengalokasikan Data ke Dalam Cluster..... | 30 |
| 4.1.3.4 | Menentukan Titik Pusat Cluster Baru..... | 31 |
| 4.1.3.5 | Memverifikasi Titik Pusat Cluster | 32 |
| 4.1.3.6 | Hasil Penerapan K-Means Clustering | 32 |
| 4.2 | Implementasi Sistem Aplikasi K-Means Clustering | 35 |
| 4.2.1 | Pembuatan Database..... | 35 |
| 4.2.2 | Pembuatan Tabel | 35 |
| 4.2.2.1 | Tabel Admin | 35 |
| 4.2.2.2 | Tabel Centroid | 36 |
| 4.2.2.3 | Tabel Objek..... | 36 |
| 4.2.2.4 | Tabel Satuan..... | 36 |
| 4.2.3 | Pembuatan Koneksi Database | 37 |
| 4.2.4 | Implementasi Interface | 37 |
| 4.2.4.1 | Tampilan Halaman Login | 38 |
| 4.2.4.2 | Tampilan Halaman Home | 38 |
| 4.2.4.3 | Tampilan Halaman Data | 38 |

| | | |
|----------------------|---|----|
| 4.2.4.4 | Tampilan Halaman Proses Clustering..... | 39 |
| 4.3 | Evaluasi Clustering Menggunakan Davies-Bouldin Index..... | 41 |
| 4.3.1 | Sum of Square Within-cluster | 41 |
| 4.3.2 | Sum of Square Between-cluster | 42 |
| 4.3.3 | Ratio | 42 |
| 4.3.4 | Davies-Bouldin Index..... | 43 |
| BAB V PENUTUP | | 48 |
| 5.1 | Kesimpulan | 48 |
| 5.2 | Saran | 49 |
| DAFTAR PUSTAKA | | 50 |



DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Simbol ERD | 11 |
| Tabel 2.2 Notasi Dasar DFD | 12 |
| Tabel 3.1 Tabel Admin | 21 |
| Tabel 3.2 Tabel Centroid | 21 |
| Tabel 3.3 Tabel Objek | 22 |
| Tabel 3.4 Tabel Satuan | 22 |
| Tabel 4.1 Titik Pusat Cluster Awal | 28 |
| Tabel 4.2 Titik Pusat Cluster Baru (Cluster 1) | 31 |
| Tabel 4.3 Titik Pusat Cluster Baru (Cluster 2) | 31 |
| Tabel 4.4 Titik Pusat Cluster Baru (Cluster 3) | 31 |
| Tabel 4.5 Tabel Hasil Perhitungan SSW | 42 |
| Tabel 4.6 Tabel Hasil Perhitungan SSB | 42 |
| Tabel 4.7 Tabel Hasil Perhitungan Rasio | 42 |
| Tabel 4.8 Data Training | 43 |
| Tabel 4.9 Data Testing | 45 |
| Tabel 4.10 Perbandingan Hasil Data | 45 |
| Tabel 4.11 Evaluasi | 47 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3.1 DFD Level 0..... | 19 |
| Gambar 3.2 DFD Level 1..... | 19 |
| Gambar 3.3 Rancangan ERD..... | 20 |
| Gambar 3.4 Rancangan Halaman Login..... | 23 |
| Gambar 3.5 Rancangan Halaman Home..... | 24 |
| Gambar 3.6 Rancangan Halaman Data..... | 25 |
| Gambar 3.7 Rancangan Halaman Proses Clustering..... | 26 |
| Gambar 4.1 Hasil Preprocessing Data..... | 27 |
| Gambar 4.2 Hasil Perhitungan Jarak ke Setiap Cluster Clustering..... | 29 |
| Gambar 4.3 Alokasi Data Cluster 1..... | 30 |
| Gambar 4.4 Verifikasi Data Centroid..... | 32 |
| Gambar 4.5 Tabel Cluster 1 Hasil Penerapan Algoritma K-Means Clustering (Microsoft Excel)..... | 33 |
| Gambar 4.6 Tabel Cluster 2 Hasil Penerapan Algoritma K-Means Clustering (Microsoft Excel)..... | 33 |
| Gambar 4.7 Tabel Cluster 3 Hasil Penerapan Algoritma K-Means Clustering (Microsoft Excel)..... | 34 |
| Gambar 4.8 Pembuatan Database..... | 35 |
| Gambar 4.9 Tabel Admin..... | 35 |
| Gambar 4.10 Tabel Centroid..... | 36 |
| Gambar 4.11 Tabel Objek..... | 36 |
| Gambar 4.12 Tabel Satukan..... | 36 |
| Gambar 4.13 Pembuatan Koneksi Database..... | 37 |
| Gambar 4.14 Tampilan Halaman Login..... | 37 |
| Gambar 4.15 Tampilan Halaman Home..... | 38 |
| Gambar 4.16 Tampilan Halaman Data..... | 38 |
| Gambar 4.17 Tampilan Halaman Proses Clustering..... | 39 |
| Gambar 4.18 Halaman Hasil Proses K-Means..... | 40 |

INTISARI

Covid-19 atau Coronavirus merupakan virus yang ditemukan di manusia dan hewan. Virus ini dapat menginfeksi manusia hingga menyebabkan berbagai penyakit seperti flu, hingga penyakit serius seperti *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). Di Indonesia penyebaran kasus Covid-19 terus bertambah dan merata di seluruh provinsi di Indonesia karena penyebaran yang cukup cepat karena luasnya wilayah di Indonesia sehingga memungkinkan diperlukannya pengelompokan berdasarkan wilayah di Indonesia yang akan menghasilkan titik-titik pusat penyebaran dari kasus Covid-19 ini.

Penelitian ini bertujuan mengelompokkan data Covid-19 kedalam sebuah cluster dengan menggunakan Data Mining Algoritma K-Means Clustering. Data Covid-19 yang digunakan dalam penelitian ini adalah data Covid-19 tanggal 6 Juli 2021 yang diambil dari website resmi Kawal Covid-19 (kawalCovid-19.id). Untuk atribut yang digunakan adalah kasus positif, sembuh, dan meninggal.

Cluster yang terbentuk dari hasil penelitian menggunakan K-Means Clustering adalah 3 cluster dengan cluster pertama berjumlah 2 provinsi, cluster kedua 3 provinsi, dan untuk cluster ketiga adalah 29 provinsi. Cluster dengan tingkat penyebaran Covid-19 terbesar adalah cluster satu. Dari penelitian ini didapatkan akurasi 91,176% dan dievaluasi menggunakan Davies-Bouldin Index menghasilkan hasil cluster cukup baik dengan nilai 0,493371469.

Kata Kunci: Covid-19, Data Mining, K-Means Clustering, Indonesia

ABSTRACT

Covid-19 or Coronavirus is a virus that is found in humans and animals. This virus can infect humans to cause various diseases such as flu, to serious diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). In Indonesia, the spread of Covid-19 cases continues to increase and is evenly distributed in all provinces in Indonesia because of the fairly rapid spread due to the vast area in Indonesia, making it possible for grouping based on regions in Indonesia to be needed which will result in the center points of the spread of this Covid-19 case.

This study aims to group Covid-19 data into a cluster using the K-Means Clustering Data Mining Algorithm. The Covid-19 data used in this study is Covid-19 data on July 6, 2021 which was taken from the official website of Kawal Covid-19 (KawalCovid-19.id). The attributes used are positive cases, recovered, and died.

The clusters formed from the results of research using K-Means Clustering are 3 clusters with the first cluster consisting of 2 provinces, the second cluster 3 provinces, and for the third cluster 29 provinces. The cluster with the largest Covid-19 spread rate is cluster one. From this study, the accuracy was 91.176% and evaluated using the Davies-Bouldin Index yielded a fairly good cluster result with a value of 0.493371469.

Keywords: Covid-19, Data Mining, K-Means Clustering, Indonesia