



**ANALISIS DAN PERBANDINGAN KELOMPOK PASIEN COVID-19
BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING***

SKRIPSI

HILMA FITRI SOLEHAH

1810512052

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2021



**ANALISIS DAN PERBANDINGAN KELOMPOK PASIEN COVID-19
BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS CLUSTERING***

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar

Sarjana Komputer

HILMA FITRI SOLEHAH

1810512052

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2021

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar

Nama : Hilma Fitri Solehah

NIM : 1810512052

Tanggal : 24 November 2021

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan hukum yang berlaku.

Jakarta, 24 November 2021

Yang Menyatakan,



Hilma Fitri Solehah

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
PENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, Saya yang bertanda tangan di bawah ini:

Nama : Hilma Fitri Solehah
NIM : 1810512052
Fakultas : Ilmu Komputer
Program Studi : S1 -Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**ANALISIS DAN PERBANDINGAN KELOMPOK PASIEN COVID-19
BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS*
*CLUSTERING***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 24 November 2021

Yang Menyatakan,



Hilma Fitri Solehah

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Hilma Fitri Solehah

NIM : 1810512052

Program Studi : S-1 Sistem Informasi

Judul Skripsi : Analisis dan Perbandingan Kelompok Pasien COVID-19 Berdasarkan Komorbiditas Menggunakan *K-Means Clustering*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Iin Ernawati, S.Kom., M.Si.

Ketua Penguji



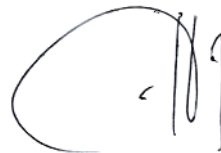
Helena Nurramdhani, S.Pd., M.Kom.

Penguji II



Ati Zaidiah, S.Kom., MTI.

Pembimbing I



Ika Nurlaili Isnainivah, S.Kom., M.Sc.

Pembimbing II



Dr. Ermatita, M.Kom.

Dekan Fakultas Ilmu Komputer



Ati Zaidiah, S.Kom., MTI.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 17 Desember 2021



**ANALISIS DAN PERBANDINGAN KELOMPOK PASIEN COVID-19
BERDASARKAN KOMORBIDITAS MENGGUNAKAN *K-MEANS*
*CLUSTERING***

Hilma Fitri Solehah

ABSTRAK

Pandemi COVID-19 yang terjadi sejak awal tahun 2020 merupakan sebuah permasalahan dan tantangan baru yang dirasakan oleh dunia. Penularan infeksi virus SARS-CoV-2 sangat mudah dan terjadi dengan cepat. Salah satu faktor risiko yang rentan terinfeksi COVID-19 adalah faktor penyakit penyerta/ komorbiditas yang ada pada masyarakat. Untuk dapat mengidentifikasi kelompok komorbid apa saja yang banyak ditemukan pada pasien COVID-19, penelitian ini melakukan kegiatan analisis dan pemodelan menggunakan algoritma data mining *K-Means Clustering* yang akan menghasilkan cluster pasien COVID-19. Adapun atribut data yang digunakan yakni nilai tengah usia, jenis komorbiditas, serta persentase penyakit melihat ukuran populasinya. Data yang akan diolah didapatkan dari sumber valid Covid Analytics yang berisikan rangkuman data riset medis pasien COVID-19 dari berbagai Negara. Cluster yang dihasilkan berjumlah 3 cluster dengan perbedaan derajat dari ringan hingga berat. Terdapat 321 data (39%) kelompok pasien COVID-19 yang diolah termasuk pada cluster pertama dengan rata-rata usia 48 tahun dan komorbid terbanyak ditemukan yaitu penyakit diabetes. Cluster kedua terdiri dari 370 data (44%) dengan rata-rata usia 64 tahun dan penyakit diabetes merupakan komorbid terbanyak. Sedangkan untuk cluster ketiga terdiri dari 140 data (17%) dengan rata-rata usia 68 tahun dan komorbid terbanyak yaitu penyakit Hipertensi. Analisis menggunakan *K-Means Clustering* ini dapat memberikan gambaran kelompok pasien dengan derajat yang berbeda serta bentuk penanganan yang akan berbeda pula.

Kata Kunci: Covid-19, Komorbiditas, *K-Means Clustering*

ANALYSIS AND COMPARISON OF COVID-19 PATIENTS BASED ON COMORBIDITY USING K-MEANS CLUSTERING

Hilma Fitri Solehah

ABSTRACT

The COVID-19 pandemic that has occurred since the beginning of 2020 is a new problem and challenge that is felt by all around the world. The transmission of the SARS-CoV-2 virus infection is very easy and fast to spread among us. Comorbidity is one of the risk factors that is susceptible to the virus infection. To identify any comorbid groups that can be found in COVID-19 patients, an analysis study using data mining method with K-Means Clustering algorithm will help to do grouping by the clustering results from the patients based on the proportion of comorbidities and age ranges that found in several populations. The data to be processed is obtained from a valid source Covid Analytics which contains a summary of medical research data for COVID-19 patients from various countries. The cluster results are 3 clusters with different clinical degrees from mild to severe. There are 321 data (39%) of the COVID-19 patients group who were treated in the first cluster with an average age of 48 years old and the most found comorbidity was diabetes. The second cluster consisted of 370 data (44%) with an average age of 64 years old with diabetes as the most found comorbidity. Meanwhile, the third cluster consisted of 140 data (17%) with an average age of 68 years and the highest comorbidity was hypertension. This analysis using K-Means Clustering can provide an overview of different groups of patients also with different clinical degrees and different forms of treatment.

Keywords: Covid-19, Comorbidities, K-Means Clustering

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Segala puji hanya bagi Allah Tuhan semesta alam, puji syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat dan karunia yang diberikanNya. Shalawat dan salam tercurahkan kepada baginda Rasulullah ﷺ beserta seluruh keluarganya, para sahabatnya, dan para pengikutnya. Alhamdulillah waa syukurillah, atas keberkahan dan kemudahan yang Allah SWT berikan, penulis dapat menyelesaikan tugas akhir ini yang berjudul **“Analisis Dan Perbandingan Kelompok Pasien Covid-19 Berdasarkan Komorbiditas Menggunakan *K-Means Clustering*”**

Penulisan Tugas Akhir ini merupakan pemenuhan syarat untuk memperoleh gelar Sarjana Ilmu Komputer (S.Kom) bagi mahasiswa program studi S1 Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta. Rasa terima kasih penulis ucapkan kepada:

1. Kedua orang tua penulis, Abi dan Umi yang memberikan do'a, kasih sayang, hadiah, semangat dan dukungan moril materil agar penulis semangat menyelesaikan tugas akhir ini tepat waktu.
2. Sanak saudara penulis, Kakak Aisyah, Abang Ali, Kakak Sheila, Husna, dan Humairah yang meramaikan dan menghebohkan rumah setiap harinya. Serta para sepupu penulis yang selalu memberikan dukungan serta keceriaan unlimited.
3. Ibu Ati Zaidiah, S.Kom., MTI., selaku Ketua Program Studi S-1 Sistem Informasi dan Dosen Pembimbing I yang berkenan memberikan masukan, bantuan, menyisihkan waktunya, dan memberikan tambahan ilmu yang bermanfaat untuk membantu penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya.
4. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc., selaku Dosen Pembimbing II yang memberikan waktunya, bimbingan ilmu, pengarahan, dan solusi yang baik kepada penulis dalam menyelesaikan skripsi ini.
5. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
6. Seluruh Bapak dan Ibu Dosen Prodi Sistem Informasi UPN Veteran Jakarta yang telah membimbing dan memberikan ilmu yang bermanfaat selama masa perkuliahan
7. Para sahabat dan teman-teman penulis, teman-teman Sumtenive yang selalu mendengarkan di tengah kepenatan, teman-teman KKC moodbooster yang selalu menghibur dan melawak, teman-teman TDP yang selalu sama-sama dan supportive dari

awal perkuliahan, May Allah make this easy for us, dan teman seperjuangan angkatan 2018 lainnya yang selalu memberikan bantuan dan sharing-sharing di kala susah dan senang, terima kasih sudah hadir menemani kisah perkuliahanku teman-teman!

8. Teman satu bimbingan penulis, Dinda dan Devina yang bersedia struggling bersama di topik data mining ini. Selamat buat kita sudah melewatinya dengan sabar sampai selesai.
9. Semua pihak lainnya yang terlibat dalam proses penulisan tugas akhir ini yang tidak dapat disebutkan satu per satu namanya.

Semoga Allah SWT membalas kebaikan semua pihak yang membantu memberikan bantuan, perhatian, do'a, dan dukungan kepada penulis dalam menyelesaikan Tugas Akhir ini. Akhir kata, penulis berharap semoga tugas akhir ini dapat memberikan kebermanfaatan ilmu bagi pembaca dan orang-orang yang membutuhkannya. Terima Kasih

Jakarta, 24 November 2021

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK PENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
DAFTAR SIMBOL	xvii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Luaran yang diharapkan	4
1.7 Sistematika Penulisan	4
BAB II	6
LANDASAN TEORI	6
2.1 COVID-19	6
2.2 Penyakit Hipertensi	7
2.3 Penyakit Diabetes	7
2.4 Penyakit Jantung Koroner (<i>Coronary Artery Disease (CAD)</i>)	8
2.5 Penyakit Paru Obstruktif Kronik (<i>Chronic Obstructive Pulmonary Disease (COPD)</i>) ..	8
2.6 Penyakit Gagal Ginjal Kronis (<i>Chronic Kidney/ Renal Disease</i>)	8
2.7 COVID-19 dan Komorbiditas	9
2.8 Derajat Penyakit COVID-19	10
2.9 Tatalaksana Pasien Terkonfirmasi COVID-19	10
2.10 Metode CRISP-DM (<i>Cross Industry Standard Process For Data Mining</i>)	11
2.10.1 <i>Business Understanding</i>	12
2.10.2 <i>Data Understanding</i>	12

2.10.3 <i>Data Preparation / Data Pre-processing</i>	12
2.10.4 <i>Modelling</i>	13
2.10.5 <i>Evaluation</i>	13
2.10.6 <i>Deployment</i>	13
2.11 Data Mining	13
2.12 Analisis Kluster	14
2.13 <i>K-Means Clustering</i>	14
2.14 Metode Elbow	15
2.15 Mengukur Evaluasi	16
2.16 Penelitian Terdahulu	17
BAB III	19
METODOLOGI PENELITIAN.....	19
3.1 Alur Penelitian	19
3.1.1 Identifikasi Masalah	19
3.1.2 Studi Literatur	19
3.1.3 Penerapan Metode CRISP-DM.....	20
3.1.3.1 Pemahaman Bisnis (<i>Business Understanding</i>).....	20
3.1.3.2 Pemahaman Data (<i>Data Understanding</i>)	21
3.1.3.3 Persiapan Data (<i>Data Pre-processing</i>).....	22
3.1.3.4 Pemodelan Data (<i>Data Modelling</i>).....	22
3.1.3.5 Evaluasi (<i>Evaluation</i>)	24
3.1.3.6 <i>Deployment</i>	24
3.3 Alat Bantu Penelitian	24
3.4 Jadwal Penelitian	25
BAB IV	26
HASIL DAN PEMBAHASAN.....	26
4.1 Pemahaman Data.....	26
4.2 Data Pre-Processing	27
4.3 Proses Pemodelan <i>K-Means Clustering</i>	32
4.3.1 Menghitung jumlah <i>cluster</i> k	32
4.3.2 Menentukan <i>centroid</i>	33
4.3.3 Hitung jarak setiap data ke pusat <i>cluster</i>	34
4.3.4 Hitung <i>centroid</i> baru	36
4.3.5 Melakukan iterasi dengan <i>centroid</i> baru sampai <i>cluster</i> tidak berubah.....	37
4.3.6 Hasil Perhitungan <i>Centroid</i> Terakhir	37
4.4 Perhitungan Evaluasi.....	54
4.5 Implementasi Sistem	54

4.5.1 <i>Use Case Diagram</i>	55
4.5.2 <i>Class Diagram</i>	56
4.5.3 Rancangan Basis Data.....	56
4.5.3.1 Tabel Atribut Data <i>Pre-processing</i>	56
4.5.3.2 Tabel Atribut Data <i>Cluster</i>	57
4.5.4 Struktur Menu.....	58
4.5.5 <i>Flowchart</i> Fitur Prediksi.....	59
4.5.6 Tampilan Sistem.....	60
4.5.6.1 Halaman <i>Dashboard</i>	60
4.5.6.2 Halaman Data <i>Pre-processing</i>	60
4.5.6.3 Halaman Data <i>Clustering</i>	61
4.5.6.4 Halaman <i>Cluster Prediction</i>	63
4.5.7 Pengujian Sistem.....	67
BAB V.....	70
PENUTUP.....	70
5.1 Kesimpulan.....	70
5.2 Saran.....	71
DAFTAR PUSTAKA.....	72
RIWAYAT HIDUP.....	75
LAMPIRAN.....	77

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu	17
Tabel 2. Jadwal Penelitian	25
Tabel 3. Sampel Dataset	26
Tabel 4. Sampel Variabel Data	27
Tabel 5. Sampel Data Penyakit Sebelum Transformasi.....	28
Tabel 6. Sampel Data Penyakit Setelah Transformasi.....	28
Tabel 7. Inisialisasi Komorbiditas	29
Tabel 8. Inisialisasi Negara.....	29
Tabel 9. Inisialisasi Jumlah Sampel/Populasi	30
Tabel 10. Keterangan Atribut Data.....	30
Tabel 11. Frekuensi Data berdasarkan Komorbid	31
Tabel 12. Data Awal Penelitian	31
Tabel 13. <i>Centroid</i> Awal.....	34
Tabel 14. Tabel Inisialisasi Perhitungan	34
Tabel 15. Tabel Hasil Iterasi pertama	35
Tabel 16. Tabel <i>Centroid</i> Baru	36
Tabel 17. Hasil <i>Centroid</i> Terakhir	37
Tabel 18. Hasil <i>Clustering</i>	37
Tabel 19. Hasil <i>Cluster</i> Pertama	38
Tabel 20. Hasil <i>Cluster</i> Kedua.....	38
Tabel 21. Hasil <i>Clustering</i> Ketiga	39
Tabel 22. Frekuensi Atribut Persentase pada <i>Cluster</i> 1	41
Tabel 23. Frekuensi Atribut Nama Negara pada <i>Cluster</i> 1.....	42
Tabel 24. Frekuensi Atribut Jumlah Sampel pada <i>Cluster</i> 1	42
Tabel 25. Frekuensi Atribut Persentase pada <i>Cluster</i> 2	44
Tabel 26. Frekuensi Atribut Nama Negara pada <i>Cluster</i> 2.....	45
Tabel 27. Frekuensi Atribut Jumlah Sampel pada <i>Cluster</i> 2	46
Tabel 28. Frekuensi Atribut Persentase pada <i>Cluster</i> 3	47
Tabel 29. Frekuensi Atribut Nama Negara pada <i>Cluster</i> 3.....	48
Tabel 30. Frekuensi Atribut Jumlah Sampel pada <i>Cluster</i> 3	48
Tabel 31. Nilai SSE pada tiap <i>cluster</i>	54
Tabel 32. Tabel Atribut Data <i>Pre-processing</i>	57
Tabel 33. Tabel Atribut Data <i>Cluster</i>	57
Tabel 34. Tabel Pengujian Sistem	67

DAFTAR GAMBAR

Gambar 1. Fase CRISP-DM	12
Gambar 2. Alur Penelitian	19
Gambar 3. Metode CRISP-DM.....	20
Gambar 4. Flowchart K-Means Clustering	23
Gambar 5. Grafik metode <i>elbow</i>	33
Gambar 6. Histogram Median Age pada Cluster 1	39
Gambar 7. Histogram <i>Comorbidities</i> pada Cluster 1	40
Gambar 8. Histogram Median Age pada Cluster 2	43
Gambar 9. Histogram <i>Comorbidities</i> pada Cluster 2	44
Gambar 10. Histogram Median Age pada <i>Cluster 3</i>	46
Gambar 11. Histogram <i>Comorbidities</i> pada Cluster 3	47
Gambar 12. Plot Hasil <i>Clustering</i>	49
Gambar 13. Prediktor COVID-19 Berderajat Berat	52
Gambar 14. <i>Use Case</i> Diagram	55
Gambar 15. <i>Class</i> Diagram.....	56
Gambar 16. Struktur Menu Sistem	58
Gambar 17. <i>Flowchart</i> Prediksi <i>Cluster</i>	59
Gambar 18. Halaman <i>Dashboard</i>	60
Gambar 19. Halaman Data <i>Pre-Processing</i>	61
Gambar 20. Halaman Data <i>Clustering</i>	62
Gambar 21. Halaman Hasil <i>Clustering</i>	63
Gambar 22. Halaman <i>Cluster Prediction</i>	64
Gambar 23. Halaman Hasil Prediksi <i>Cluster 1</i>	64
Gambar 24. Halaman Hasil Prediksi <i>Cluster 2</i>	65
Gambar 25. Halaman Hasil Prediksi <i>Cluster 3</i>	65
Gambar 26. Halaman Hasil Prediksi <i>Cluster 1 dan 2</i>	66

DAFTAR RUMUS

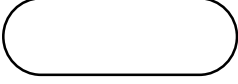


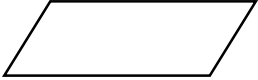
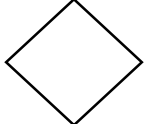
(2.1) Rumus <i>Euclidean Distance</i>	15
(2.2) Rumus <i>Sum of Square Error (SSE)</i>	15

DAFTAR LAMPIRAN

Lampiran 1 Histogram Atribut <i>Cluster 1</i>	78
Lampiran 2 Histogram Atribut <i>Cluster 2</i>	79
Lampiran 3 Histogram Atribut <i>Cluster 3</i>	80
Lampiran 4 Syntax Python <i>Pre-Processing Data</i>	81
Lampiran 5 Syntax Python <i>Modelling Data</i>	84
Lampiran 6 <i>Source Code</i> Sistem dengan Python Flask	89
Lampiran 7 <i>Raw Data Covid Analytics</i>	91
Lampiran 8 Perhitungan <i>Clustering Manual</i>	97
Lampiran 9 Hasil Uji Turnitin	109

DAFTAR SIMBOL

Simbol *Flow Chart* (Diagram Alir)

No	Simbol	Keterangan
1		Mulai/ Selesai (<i>Terminator</i>)
2		Aliran Data
3		Proses
4		Pernyataan yang berhubungan dengan Input/ Output
5		Percabangan (<i>Decision</i>)